

Diffusive Transport in Quasi-2D and Quasi-1D Electron Systems

I. Knezevic^a, E. B. Ramayya^a, D. Vasileska^b, and S. M. Goodnick^b

^a*Department of Electrical and Computer Engineering,*

University of Wisconsin–Madison, Madison, WI 53706, USA

^b*Department of Electrical Engineering, Fulton School of Engineering,
Arizona State University, Tempe, AZ 85287, USA*

Quantum-confined semiconductor structures are the cornerstone of modern-day electronics. Spatial confinement in these structures leads to formation of discrete low-dimensional subbands. At room temperature, carriers transfer among different states due to efficient scattering with phonons, charged impurities, surface roughness and other electrons, so transport is scattering-limited (diffusive) and well described by the Boltzmann transport equation. In this review, we present the theoretical framework used for the description and simulation of diffusive electron transport in quasi-two-dimensional and quasi-one-dimensional semiconductor structures. Transport in silicon MOSFETs and nanowires is presented in detail.

Keywords: quantum confinement, nanostructures, 2DEG, nanowires, SiNW, scattering, diffusive transport, Boltzmann transport equation, Monte Carlo simulation, confined phonons

| Contents | | B. Scattering due to Confined Acoustic Phonons | 22 |
|---|----|--|----|
| I. Introduction | 1 | VII. Transport in Quasi-1D Electron Systems | 23 |
| II. Quasi-2D Electron Systems | 2 | A. Gated Rectangular Nanowires | 24 |
| A. Silicon Inversion Layers | 2 | B. Gated Square Nanowires | 25 |
| B. 1D Schrödinger-Poisson Problem for Silicon Inversion Layers | 4 | 1. Subband Modulation | 26 |
| C. GaAs/AlGaAs Heterostructures. Effective Mass Schrödinger Equation for Heterostructures | 6 | 2. Volume Inversion | 26 |
| D. Density of States (DOS) for Low-Dimensional Systems | 7 | 3. Mobility Variation with the SiNW Cross Section | 27 |
| III. Scattering in Quasi-2D Electron Systems | 8 | C. Ungated Silicon Nanowires | 27 |
| A. Electron-phonon scattering | 9 | VIII. Concluding remarks | 28 |
| 1. Deformation potential scattering | 10 | Acknowledgment | 28 |
| 2. Nonpolar optical phonon scattering | 12 | References | 28 |
| 3. Polar optical phonon scattering | 13 | | |
| B. Scattering by Electrostatic Interactions | 13 | | |
| 1. Coulomb Scattering | 13 | | |
| 2. Surface-roughness scattering | 14 | | |
| 3. Electron-Electron Interaction | 16 | | |
| 4. Screening of Coulomb, Surface-Roughness and Electron-Electron Scattering | 16 | | |
| IV. Transport in Quasi-2D Systems | 16 | | |
| A. Mobility Calculation in Silicon Inversion Layers Using 2D Monte Carlo and Including Degeneracy Effects | 17 | | |
| V. Quasi-1D Semiconductors (Nanowires) | 18 | | |
| A. Electronic Bandstructure Modification in Nanowires | 19 | | |
| B. Acoustic Phonon Confinement in SiNWs | 20 | | |
| VI. Scattering in Quasi-1D Electron Systems | 21 | | |
| A. Scattering due to Bulk Acoustic Phonons, Intervalley Phonons, and Surface Roughness | 21 | | |

I. INTRODUCTION

Quantum-confined semiconductor structures, such as silicon MOSFETs or GaAs-based resonant-tunneling diodes, are the cornerstone of modern-day electronics.^{1,2,3} Spatial confinement in these structures leads to formation of discrete low-dimensional subbands: energy levels are quantized in each direction of confinement, while the momentum remains a good (continuous) quantum number in the unconfined directions. If carriers are confined along one direction and free to move in the two-dimensional (2D) plane perpendicular to it, the structure is being referred to as a *quasi-two-dimensional electron gas* (Q2DEG). The structure is considered a *quasi-one-dimensional electron gas* (Q1DEG) if the carrier motion is unbound in one dimension (1D) as a result of 2D-confinement. At room temperature, transport within each subband and transitions among subbands can essentially be described semiclassically, using the Boltzmann transport equation. Its state-of-the-art

solution is obtained via the ensemble Monte Carlo technique.⁴ At low temperatures, quantum-coherence effects become prominent, and transport ceases to be semiclassical in nature. In this regime, transport description is better achieved using the Wigner-function formalism⁵ or nonequilibrium Green's functions,^{6,7} as long as the single particle picture is valid. However, in the remainder of this text, we will be concerned with the semiclassical transport picture, appropriate for room-temperature operation of Q2DEG and Q1DEG electronic structures.

In structures such as the resonant tunneling diode⁸ or the quantum cascade laser,⁹ the actual device operation is based on utilization of quantum confinement and tunneling. In contrast, in structures such as silicon MOSFETs, quantum-confinement features emerge as a result of miniaturization and are usually detrimental: tunneling through the gate oxide,¹⁰ source-to-drain tunneling and space-quantization effects are expected to be important in nanoscale MOSFETs and HEMT devices and require a solution of the 1D Schrödinger-Poisson problem. Solution of the 2D Schrödinger-Poisson problem is needed, for example, for describing the channel charge in narrow-width MOSFETs¹¹ and nanowires.^{12,13}

Successful scaling of MOSFETs towards shorter channel lengths requires thinner gate oxides and higher doping levels to achieve high drive currents and minimized short-channel effects.^{14,15} As the oxide thickness is scaled to below 10 nm, quantum confinement of inversion charge leads to an appreciable inversion layer capacitance^{16,17} in series with the oxide, so the total gate capacitance is lowered. (Further modification of the gate oxide capacitance stems from the image and many-body exchange-correlation effects in the inversion layer,¹⁸ as well as polysilicon gate depletion.¹⁹)

The low-field electron mobility is an important quantity that determines the performance of semiconductor devices. Surface roughness scattering (SRS) is by far the most important cause of mobility degradation in conventional MOSFETs at high transverse fields. One would expect the SRS to be even more detrimental in silicon nanowires (SiNWs) than in conventional MOSFETs because SiNWs have four Si-SiO₂ interfaces, as opposed to one such interface in conventional MOSFETs. This has indeed been confirmed recently on ultrathin cylindrical¹² and rectangular nanowires.¹³ In addition, confinement leads to a modification in the acoustic phonon spectrum in SiNWs,^{13,20} which leads to increased electron-phonon scattering and a lowered electron mobility.^{21,22,23}

In this review, we present the theoretical framework typically used for the description and simulation of diffusive electron transport in quasi-2D and quasi-1D semiconductor structures. In Sec. II, we overview the formation of the Q2DEG in Si inversion layers and GaAs-modulation doped heterostructures. We discuss the solution to the 1D Schrödinger-Poisson problem in the direction of confinement and the density of states calculation. In Sec. III, we overview the scattering mechanisms

in semiconductors, their origin and model Hamiltonians, and give the scattering rates for Q2DEGs in the Born approximation. In Sec. IV, we present the ensemble Monte Carlo simulation of the electron mobility in the inversion layer Q2DEG formed near the Si/SiO₂ interface of a nanoscale MOSFET. Section V introduces the Q1DEG in nanowires, and overviews phonon confinement and band-structure modification in these structures. In Sec. VI, scattering rates for the Q1DEG are given, while the electron mobility in thin silicon nanowires, as obtained from detailed ensemble Monte Carlo simulations, is presented in Sec. VII.

II. QUASI-2D ELECTRON SYSTEMS

A. Silicon Inversion Layers

The best known examples of the Q2DEG are silicon MOSFETs and GaAs/AlGaAs heterostructures. An integral part of any MOSFET device is the metal-oxide-semiconductor (MOS) capacitor. Regarding the MOS capacitors, the induced interface charge is closely linked to the shape of the electron energy bands of the semiconductor near the interface. At zero applied voltage, the bending of the energy bands is ideally determined by the difference in the work functions of the metal and the semiconductor. This band bending changes with the applied bias and the bands become flat when we apply the so-called flat-band voltage given by

$$V_{FB} = \Phi_M - \Phi_{sc}, \quad (1)$$

where Φ_M and Φ_{sc} are the work functions of the metal and the semiconductor, respectively. The various energies involved are indicated in Fig. 1, where we show typical band diagrams of an MOS capacitor at zero bias. χ_s is the electron affinity for the semiconductor, E_c is the energy of the conduction band edge, and E_F is the Fermi level at zero applied voltage.

In stationary conditions, no net current flows in the direction perpendicular to the interface, owing to the very high resistance of the insulator layer (however, this does not apply to very thin oxides of a few nanometers, where tunneling becomes important). Hence, the Fermi level will remain constant inside the semiconductor, irrespective of the biasing conditions. However, between the semiconductor and the metal contact, the Fermi level is shifted by $E_{FM} - E_{Fs} = eV_G$ (see Figs. 2 and 3). Hence, we have a quasi-equilibrium situation in which the semiconductor can be treated as if in thermal equilibrium.

An MOS structure with a *p*-type semiconductor will enter the *accumulation regime* of operation when the voltage applied between the metal and the semiconductor is more negative than the flat-band voltage ($V_{FB} < 0$ in Fig. 2). In the opposite case, when $V_G > V_{FB}$, the semiconductor-oxide interface first becomes depleted of

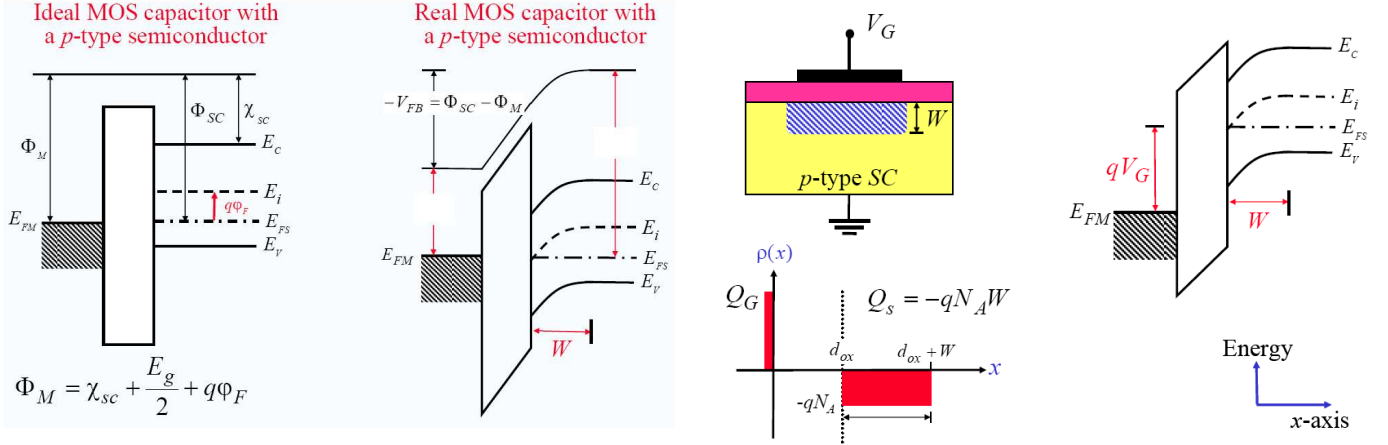


FIG. 1. (Left panel) Energy band diagram of an ideal MOS capacitor. (Right panel) Energy band diagram of a real MOS capacitor.

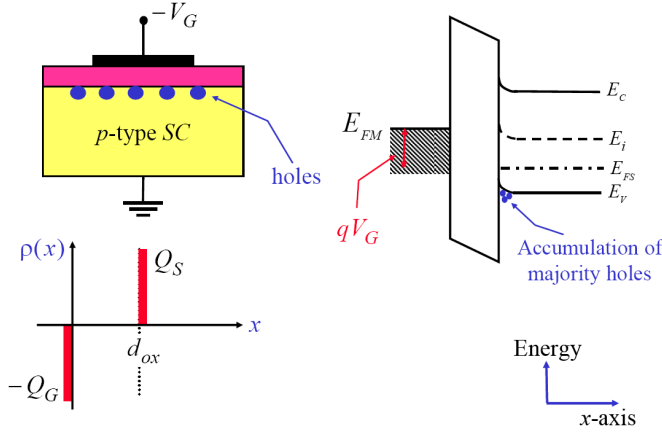


FIG. 2. MOS capacitor under accumulation conditions.

holes and we enter the so-called *depletion regime* (Fig. 3, top panel). By increasing the applied voltage, the band bending becomes so large that the energy difference between the Fermi level and the bottom of the conduction band at the insulator-semiconductor interface becomes smaller than that between the Fermi level and the top of the valence band. This is the *inversion regime* of operation (Fig. 3, bottom panel).

Carrier statistics tells us that the electron concentration will then exceed the hole concentration near the interface and we enter the inversion regime. At a larger still applied voltage, we finally arrive at a situation in which the electron density at the interface exceeds the doping density in the semiconductor. This is the strong inversion case, in which we have a significant conducting sheet of inversion charge at the interface (Fig. 3, bottom panel). In the description that follows, symbol φ is used to denote the potential in the semiconductor measured relative to the potential at a position x deep inside the semiconductor. Note that φ becomes positive when the

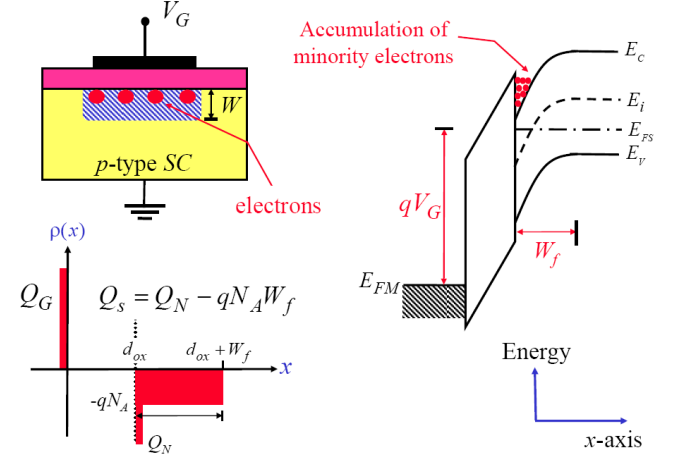


FIG. 3. MOS capacitor under depletion conditions (top panel) and inversion conditions (bottom panel).

bands bend down, as in the example of a *p*-type semiconductor shown in Figure 3. From equilibrium electron statistics, we find that the intrinsic Fermi level E_i in the bulk corresponds to an energy separation $e\varphi_F$ from the actual Fermi level E_F of the doped semiconductor,

$$\varphi_F = V_T \ln \left(\frac{N_A}{n_i} \right) > 0, \quad (2)$$

where V_T is the thermal voltage, N_A is the shallow acceptor density in the *p*-type semiconductor and n_i is the intrinsic carrier density of silicon. According to the usual definition, strong inversion is reached when the total band bending equals $2e\varphi_F$, corresponding to the surface potential $\varphi_s = 2\varphi_F$. Values of the surface potential such that $0 < \varphi_s < 2\varphi_F$ correspond to the depletion and the weak inversion regimes, respectively, $\varphi_s = 0$ is the flat-band condition, and $\varphi_s < 0$ corresponds to the accumulation mode. Note that, deep inside the semiconductor, we have $\varphi(\infty) = 0$. Under the flat-band condition ($V_G = V_{FB}$), the surface charge is equal to zero. In accumulation ($V_G < V_{FB}$), the surface charge is positive,

and in depletion and inversion ($V_G > V_{FB}$), the surface charge is negative.

In order to relate the semiconductor surface potential φ_s to the applied voltage V_G , we have to investigate how this voltage is divided between the insulator and the semiconductor. Using the condition of continuity of the electric flux density at the semiconductor-insulator interface, we find $\epsilon_s F_s = \epsilon_{ox} F_{ox}$, where ϵ_{ox} and ϵ_s are the absolute permittivities (also known as dielectric constants) of the oxide layer and the semiconductor, respectively, while F_{ox} and F_s are the respective electric fields in the two materials. Hence, for an insulator of thickness d_{ox} , the voltage drop across the insulator becomes $F_{ox} d_{ox}$. Accounting for the flat-band voltage, the applied voltage can be written as $V_G = V_{FB} + \varphi_s + \epsilon_s F_s / C_{ox}$, where $C_{ox} = \epsilon_{ox} / d_{ox}$ is the insulator capacitance per unit area.

The threshold voltage, V_{th} , is the gate voltage corresponding to the onset of strong inversion. It is one of the most important parameters characterizing metal-insulator-semiconductor devices. As discussed above, strong inversion occurs when the surface potential, φ_s , becomes equal to $2\varphi_F$. For this surface potential, the free charge induced at the insulator-semiconductor interface is still small compared to the charge in the depletion layer, and the threshold voltage is calculated using:

$$V_G = \varphi_s + \frac{\sqrt{2qN_A\epsilon_s\varphi_s}}{C_{ox}}, \text{ where } C_{ox} = \frac{\epsilon_{ox}}{d_{ox}},$$

$$V_{th} = V_G \text{ for which } \varphi_s = 2\varphi_F. \quad (3)$$

Note that the threshold voltage may also be affected by the so-called fast surface states at the semiconductor-oxide interface and by fixed charges in the insulator layer. However, this is not a significant concern with modern-day fabrication technology.

The threshold voltage separates the subthreshold regime, where the mobile carrier charge increases exponentially with increasing applied voltage V_G , from the above-threshold regime, where the mobile carrier charge is linearly dependent on the applied voltage V_G . However, there is no clear point of transition between the two regimes, so different definitions and experimental techniques have been used to determine V_{th} . Well above threshold, the charge density of the mobile carriers in the inversion layer can be calculated using the parallel-plate charge control model. This model gives an adequate description for the strong inversion regime of the MOS capacitor, but fails for applied voltages near and below threshold (i.e., in the weak inversion and depletion regimes). Several expressions have been proposed for a unified charge control model (UCCM) that covers all regimes of operation.

Successful scaling of MOSFETs towards shorter channel lengths requires thinner gate oxides and higher doping levels to achieve high drive currents and minimized short-channel effects.^{14,15} For these nanometer devices it was demonstrated that, as the oxide thickness is scaled to 10 nm and below, the total gate capacitance is lower

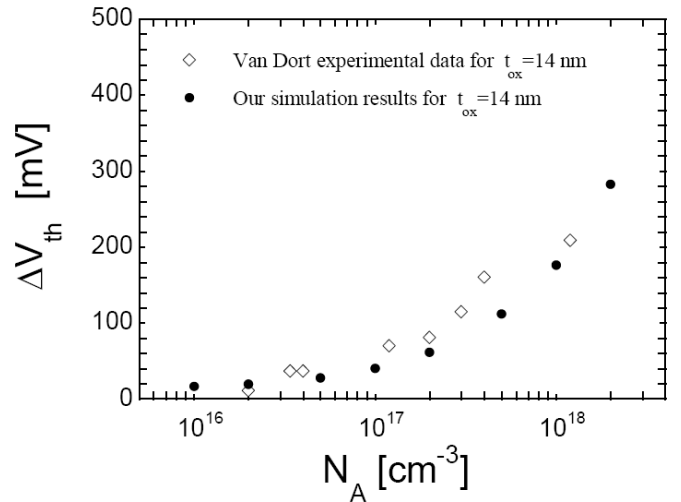


FIG. 4. SCHRED³⁰ simulation data for the shift in the threshold voltage compared to the experimental values provided by van Dort and co-workers.^{27,28}

than the oxide capacitance due to the comparable values of the oxide and the inversion layer capacitances. As a consequence, the device transconductance is degraded relative to the expectations of the scaling theory.²⁴

The two physical origins of the inversion layer capacitance, the finite density of states and the finite inversion layer thickness, were demonstrated experimentally by Takagi and Toriumi.¹⁶ A computationally efficient three-subband model, that predicts both the quantum-mechanical effects in the electron inversion layers and the electron distribution within the inversion layer, was proposed and implemented into the PISCES simulator.¹⁷ The influence of the image and many-body exchange-correlation effects on the inversion layer and the total gate capacitance was studied by Vasileska *et al.*¹⁸ It was also pointed out that the depletion of the poly-silicon gates considerably affects the magnitude of the total gate capacitance.¹⁹

The inversion layer capacitance was also identified as being the main cause of the second-order thickness dependence of the MOSFET IV-characteristics.²⁵ The finite inversion layer thickness was estimated experimentally by Hartstein and Albert.²⁶ The high levels of substrate doping were found responsible for the increased threshold voltage and decreased channel mobility. A simple analytical model that accounts for this effect was proposed by van Dort and co-workers^{27,28} and confirmed by Vasileska and Ferry²⁹ by investigating the doping dependence of the threshold voltage in MOS capacitors (Fig. 4).

B. 1D Schrödinger-Poisson Problem for Silicon Inversion Layers

The periodic crystal potential in bulk semiconducting materials is such that, for a given energy in the conduc-

tion band, the allowed electron wavevectors trace out a surface in \mathbf{k} -space. In the effective-mass approximation for silicon, these constant energy surfaces can be visualized as six equivalent ellipsoids of revolution (Fig. 5), whose major and minor axes are inversely proportional to the effective masses. A collection of such ellipsoids for different energies is referred to as a valley.

In this framework, the bulk Hamiltonian for an electron, residing in one of these valleys is of the form

$$H_0(\mathbf{R}) = -\left(\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2} + \frac{\hbar^2}{2m_y^*}\frac{\partial^2}{\partial y^2} + \frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2}\right) + V_{\text{eff}}(z) \\ = H_{0\parallel}(\mathbf{r}) + H_{0\perp}(z), \quad (4)$$

where $\mathbf{R} = (\mathbf{r}, z)$, $V_{\text{eff}}(z) = V_H(z) + V_{\text{exc}}(z)$ is the effective potential energy profile of the confining potential along the z -direction, $V_H(z)$ is the Hartree potential which is nothing more but a solution of the 1D Poisson equation introduced later in the text, $V_{\text{exc}}(z)$ is the exchange-correlation potential (also discussed later in the text), $H_{0\parallel}$ is the parallel part of H_0 (associated with the motion in the xy -plane, perpendicular to the confinement direction), and the transverse part is defined as

$$H_{0\perp}(z) = -\frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2} + V_{\text{eff}}(z). \quad (5)$$

An accurate description of the charge in the inversion layer of deep-submicrometer devices and, therefore, the magnitude of the total gate capacitance, C_{tot} , requires a self-consistent solution of the 1D Poisson equation

$$\frac{\partial}{\partial z} \left[\epsilon(z) \frac{\partial \varphi}{\partial z} \right] = -e [N_D^+(z) - N_A^-(z) + p(z) - n(z)], \quad (9)$$

and the 1D Schrödinger equation

$$\left[-\frac{\hbar^2}{2m_v^\perp}\frac{\partial^2}{\partial z^2} + V_{\text{eff}}(z) \right] \psi_{vj}(z) = \mathcal{E}_{vj} \psi_{vj}(z). \quad (10)$$

In Eqs. (9) and (10), $\varphi(z)$ is the electrostatic potential [the Hartree potential $V_H(z) = -e\varphi(z)$], $\epsilon(z)$ is the spatially dependent dielectric constant, N_D^+ and N_A^- are the ionized donor and acceptor concentrations, $n(z)$ and $p(z)$ are the electron and hole densities, V_{eff} is the effective potential energy term that equals the sum of the Hartree and exchange-correlation corrections to the ground state energy of the system, m_v^\perp is the effective mass normal to the semiconductor-oxide interface of the v -th valley, while \mathcal{E}_{vj} and $\psi_{vj}(z)$ are the energy level and the corresponding wavefunction of the electrons residing in the j -th subband from the v -th valley. The electron-density is calculated using

$$n(z) = \sum_{v,j} N_{vj} |\psi_{vj}(z)|^2, \quad (11)$$

The basis-states of the unperturbed Hamiltonian are assumed to be of the form

$$\psi_n(\mathbf{R}) = \frac{1}{\sqrt{A}} e^{i\mathbf{k}\cdot\mathbf{r}} \psi_n(z), \quad (6)$$

where \mathbf{k} is a wavevector in the xy -plane and A is the area of the sample interface. The subband wavefunctions satisfy the one-dimensional Schrödinger equation

$$H_{0\perp}(z) \psi_n(z) = \mathcal{E}_n \psi_n(z) \quad (7)$$

subject to the boundary conditions that $\psi_n(z)$ are zero for $z = 0$ and approach zero as $z \rightarrow \infty$. In Eq. (7), \mathcal{E}_n is the subband energy and $\psi_n(z)$ is the corresponding wavefunction. In the parabolic band approximation, the total energy of an electron is given by

$$\mathcal{E}_n(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m_{xy}^*} + \mathcal{E}_n = \mathcal{E}_{\mathbf{k}} + \mathcal{E}_n, \quad (8)$$

where $\mathcal{E}_{\mathbf{k}}$ is the kinetic energy and m_{xy}^* is the density-of-states mass along the xy -plane.

where N_{vj} is the sheet electron concentration in the j -th subband from the v -th valley, given by

$$N_{vj} = \nu_v \frac{m_{xy}^*}{\pi \hbar^2} k_B T \ln \{1 + \exp[(E_F - \mathcal{E}_{vj})/k_B T]\} \quad (12)$$

where ν_v is the valley degeneracy factor and E_F is the Fermi energy. When evaluating the exchange-correlation corrections to the chemical potential, we have relied on the validity of the density functional theory (DFT) of Hohenberg and Kohn,³¹ and Kohn and Sham.³² According to DFT, the effects of exchange and correlation can be included through a one-particle exchange-correlation term $V_{\text{exc}}[n(z)]$, defined as a functional derivative of the exchange-correlation part of the ground-state energy of the system with respect to the electron density $n(z)$. In the local density approximation (LDA), one replaces the functional $V_{\text{exc}}[n(z)]$ with a function $V_{\text{exc}}[n(z)] = \mu_{\text{exc}}[n_0 = n(z)]$, where μ_{exc} is the exchange-correlation contribution to the chemical potential of a homogeneous electron gas of density n_0 , which is taken to be equal to the local electron density $n(z)$ of the inhomogeneous system. In our model, we use the LDA and approximate the exchange-correlation potential energy term $V_{\text{exc}}[n(z)]$ by an interpolation formula developed by Hedin and Lundqvist.³³ Exchange and correlation effects tend to lower the total energy of the system, and, as discussed later, lead to a non-uniform shift of the

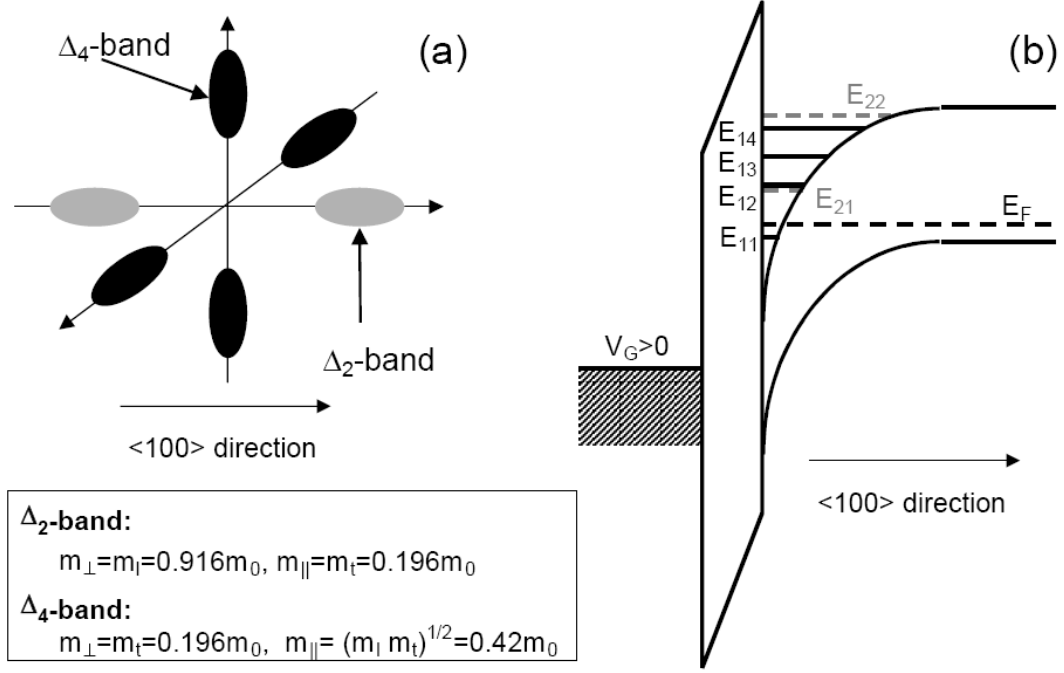


FIG. 5. (Left panel) Constant-energy surfaces for the conduction-band of silicon, showing six conduction-band valleys in the $\langle 100 \rangle$ direction of momentum space. The band minima, corresponding to the centers of the ellipsoids, are 85% of the way to the Brillouin-zone boundaries. The long axis of an ellipsoid corresponds to the longitudinal effective mass of the electrons in silicon, $m_l = 0.916m_0$, while the short axes correspond to the transverse effective mass, $m_t = 0.190m_0$ (m_0 is the free-electron mass). For a (100) surface, the Δ_2 -band has the longitudinal mass (m_l) perpendicular to the semiconductor interface and the Δ_4 -band has the transverse mass (m_t) perpendicular to the interface. Since larger mass leads to a smaller kinetic term in the Schrödinger equation, the unprimed ladder of subbands (as it is usually called), corresponding to the Δ_2 -band, has the lowest ground state energy. The degeneracy of the unprimed ladder of subbands for a (100) surface is 2. For the same reason, the ground state of the primed ladder of subbands corresponding to the Δ_4 -band is higher than the lowest subband of the unprimed ladder of subbands. The degeneracy of the primed ladder of subbands for a (100) surface is 4. (Right panel) Potential diagram for inversion of p -type silicon. The notation vj refers to the j -th subband of either the Δ_2 -band ($v = 1$) or Δ_4 -band ($v = 2$).

energy levels and repopulation of the various subbands. The enhancement of the exchange-correlation contribution to the energy predominantly affects the ground subband of the occupied valley; the unoccupied subbands of the same valley are essentially unaffected. As a result, a noticeable increase in the energy of intersubband transitions can be observed at high electron densities (more on electron scattering can be found in Sec. III).

C. GaAs/AlGaAs Heterostructures. Effective Mass Schrödinger Equation for Heterostructures

While our primary focus in this paper will remain on Si-based electron system, in this section we will briefly discuss another important Q2DEG: the modulation-doped GaAs-AlGaAs heterostructure.² The bandgap in AlGaAs is wider than in GaAs. By variation of doping it is possible to move the Fermi level inside the forbidden gap. When the materials are put together, a unified chemical potential is established, and an inversion layer is formed at the interface (see Fig. 6).

The Q2DEG created by modulation doping can be squeezed into narrow channels by selective depletion in spatially separated regions. The simplest lateral confinement technique is to create split metallic gates in a way shown in Fig. 7. We refer the reader to extensive literature on transport properties of GaAs/AlGaAs and other modulation doped heterostructures (for instance, see Ref. 3).

In semiconductors, some of the most interesting applications of the Schrödinger equation in the effective mass approximation involve spatially varying material compositions and heterojunctions. The effective mass approximation can still be used with some caution. Since the effective mass is a property of the bulk, it is not well defined in the neighborhood of a sharp material transition. In the hypothesis of slow material composition variations in space, one can adopt the Schrödinger equation with a spatially varying effective mass, taken to be the mass of the bulk with the local material properties. However, it can be shown that the Hamiltonian operator is no longer Hermitian for varying mass. A widely used Hermitian form brings the effective mass inside the differential op-

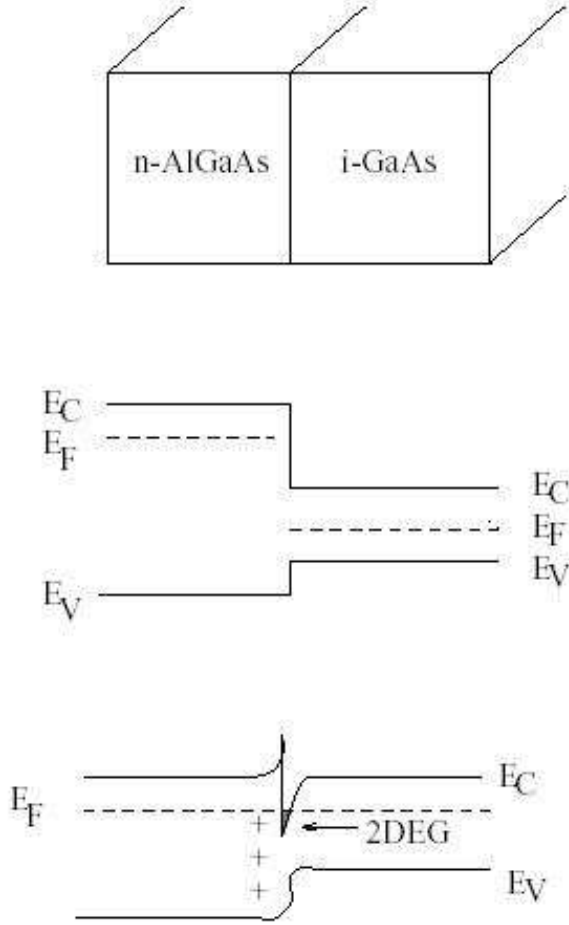


FIG. 6. Band diagrams near the interface between n-AlGaAs and intrinsic GaAs, (a) before and (b) after the charge transfer.

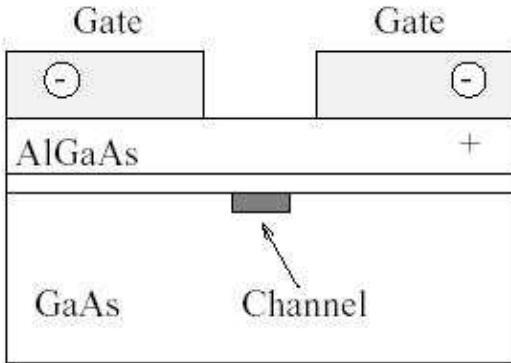


FIG. 7. On the formation of a narrow channel by a split gate.

erator as $-\frac{\hbar^2}{2}\nabla \cdot \left(\frac{1}{m^*}\nabla\psi\right)$. This approach is extended to abrupt heterojunctions, as long as the materials on the two sides have similar properties and bandstructure, as in the case of the GaAs/AlGaAs system with less than 45 % Al. One has to keep in mind that very close to the heterojunctions the effective mass Schrödinger equation provides a reasonable mathematical connection between the two regions, but the physical quantities are not necessarily well defined. For instance, in the case of a narrow potential barrier obtained by using a thin layer of AlGaAs surrounded by GaAs, it is not clear at all what effective mass should be used for the AlGaAs, since such a region can certainly not be approximated by the bulk. It is even more difficult to treat the case when there is a transition between direct and indirect bandgap materials (for example, GaAs and AlGaAs with over 45% Al).

Assuming a uniform mesh size Δx , the Hamiltonian of the Schrödinger equation can be discretized in 1D by introducing midpoints in the mesh on both sides of a generic grid point i . First, we evaluate the outer derivative at point i with centered finite differences, using quantities defined at points $(i - 1/2)$ and $(i + 1/2)$

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial x}\left[\frac{1}{m^*}\frac{\partial\psi}{\partial x}\right]_i \approx -\frac{\hbar^2}{2\Delta x}\left[\left(\frac{1}{m^*}\frac{\partial\psi}{\partial x}\right)_{i+1/2} - \left(\frac{1}{m^*}\frac{\partial\psi}{\partial x}\right)_{i-1/2}\right], \quad (13)$$

and then the derivatives defined on the midpoints are also evaluated with centered differences using quantities on the grid points

$$-\frac{\hbar^2}{2(\Delta x)^2}\left[\frac{\psi(i+1) - \psi(i)}{m^*(i+1/2)} - \frac{\psi(i) - \psi(i-1)}{m^*(i-1/2)}\right]. \quad (14)$$

The effective mass is the only quantity which must be known at the midpoints. If an abrupt heterojunction is located at point i , the abrupt change in the effective mass is treated without ambiguity.

D. Density of States (DOS) for Low-Dimensional Systems

An important quantity characterizing a quantum-mechanical system is the density of states (DOS) function. The density of states $g(\mathcal{E})$ is defined as the number of states per energy interval $(\mathcal{E}, \mathcal{E} + d\mathcal{E})$. It is clear that

$$g(\mathcal{E}) = \sum_{\alpha} \delta(\mathcal{E} - \mathcal{E}_{\alpha}), \quad (15)$$

where α is the set of quantum numbers characterizing the states. In the present case, it includes the subband quantum number n , spin quantum number σ , valley quantum number v and the in-plane quasi-momentum \mathbf{k} . If the spectrum is degenerate with respect to spin and valleys,

one can define the spin degeneracy ν_s and the valley degeneracy ν_v to get

$$g(\mathcal{E}) = \frac{\nu_s \nu_v}{(2\pi)^D} \sum_n \int d^D k \delta(\mathcal{E} - \mathcal{E}_n). \quad (16)$$

Here we calculate the number of states per unit volume, D being the dimensionality of the space. For a 2D case, we obtain

$$g(\mathcal{E}) = \frac{\nu_s \nu_v m^*}{(2\pi) \hbar^2} \sum_n \Theta(\mathcal{E} - \mathcal{E}_n). \quad (17)$$

Within a given subband, the 2D density of states function is energy-independent. Since there can exist several subbands in the confining potential, the total density of states can be represented as a set of steps, as shown in Fig. 8. At a low temperature ($k_B T \ll E_F$), all the states are filled up to the Fermi level. Because of the energy-independent density of states, the sheet electron density is linear in the Fermi energy, namely

$$N_s = N \frac{\nu_s \nu_v m^* E_F}{(2\pi) \hbar^2}. \quad (18)$$

The Fermi momentum in each subband can be determined as

$$k_{Fn} = \frac{\sqrt{2m(E_F - \mathcal{E}_n)}}{\hbar}. \quad (19)$$

In Eq. (18), N is the number of transverse modes having the edge \mathcal{E}_n below the Fermi energy.

The situation is more complicated if the gas is further confined in a narrow channel, say, along the y -axis. The in-plane wave function can be decoupled as a product

$$\psi(\mathbf{r}) \sim \eta(y) e^{ik_x x}, \quad (20)$$

the corresponding energy being

$$\mathcal{E}_{n,s,k} = \mathcal{E}_n + \mathcal{E}_s + \frac{\hbar^2 k_x^2}{2m}. \quad (21)$$

In the last equation, $\mathcal{E}_{ns} = \mathcal{E}_n + \mathcal{E}_s$ characterizes the energy level in the potential confined in both (z and y) directions. For square-box confinement, the terms are

$$\mathcal{E}_s = \frac{(s\pi\hbar)^2}{2mW^2}, \quad (22)$$

where W is the channel width, while for parabolic confinement $U(y) = (1/2)m\omega_0^2 y^2$ (typical for split-gate structures), we have

$$\mathcal{E}_s = (s - 1/2) \hbar\omega_0. \quad (23)$$

For these systems, confined in 2D, the total density of states is

$$g(\mathcal{E}) = \frac{\nu_s \nu_v \sqrt{m}}{2^{3/2} \pi \hbar} \sum_{n,s} \frac{\Theta(\mathcal{E} - \mathcal{E}_{ns})}{\sqrt{\mathcal{E} - \mathcal{E}_{ns}}}. \quad (24)$$

The energy dependence of the density of states is shown in Fig. 9.

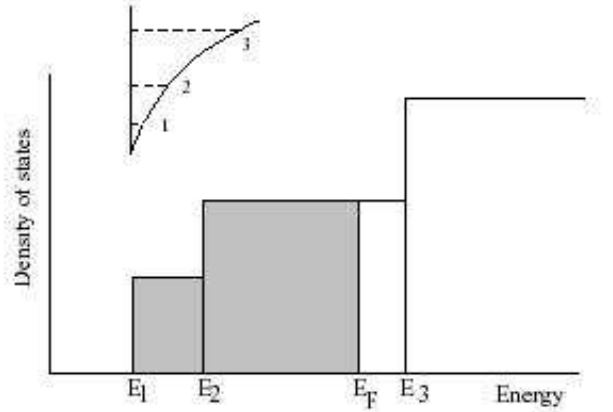


FIG. 8. Density of states for a quasi-2D system. The three lowest subbands are included.

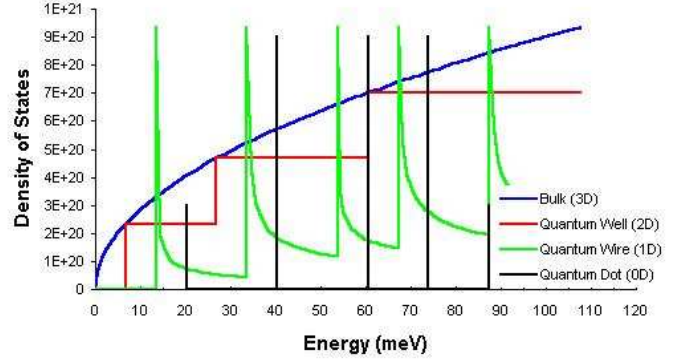


FIG. 9. Density of states for the bulk (3D, blue), quantum well (2D, red), quantum wire (1D, green) and quantum dot (0D, black).

III. SCATTERING IN QUASI-2D ELECTRON SYSTEMS

Charge transport in the diffusive regime is governed by carrier scattering from lattice vibrations, charged impurities, defects, interface roughness, as well as other electrons. Calculation of the scattering rates for confined carriers proceeds in a similar manner as in the 3D case,^{2,34} but proper wavefunctions for 2D carriers must be used. Before we go into the details of the calculation of the matrix elements of some of the most important scattering mechanisms listed in Fig. 10, we will derive a few expressions.

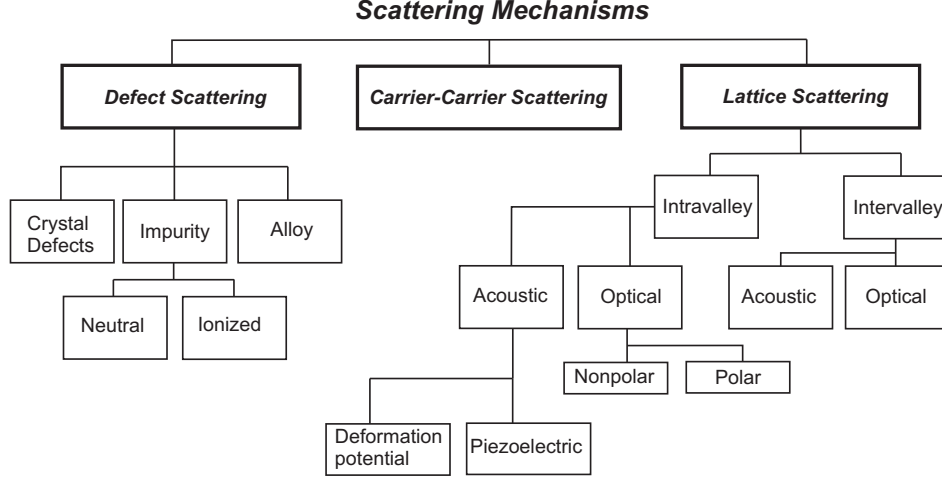


FIG. 10. Scattering mechanisms in a typical semiconductor.

Suppose we want to calculate the scattering rate out of some state \mathbf{k} in subband n . For that purpose we will use Fermi's golden rule, which gives us the transition rate from state \mathbf{k} in subband n into state \mathbf{k}' belonging to subband m by means of emission of an energetic particle (e.g. a phonon or a photon) with energy $\hbar\omega$:

$$S_{nm}(\mathbf{k}, \mathbf{k}') = \frac{2\pi}{\hbar} |M(\mathbf{k}, \mathbf{k}')|_{nm}^2 \delta(\mathcal{E}' - \mathcal{E} + \hbar\omega). \quad (25)$$

Assuming a plane-wave basis for the wavefunctions in the unconfined direction (xy -plane), the total wavefunctions of the initial and the final states are of the following general form (for a Q2DEG):

$$\psi_n(\mathbf{k}, z) = \frac{1}{\sqrt{A}} e^{i\mathbf{k}\cdot\mathbf{r}} \varphi_n(z), \quad \psi_m(\mathbf{k}', z') = \frac{1}{\sqrt{A}} e^{i\mathbf{k}'\cdot\mathbf{r}} \varphi_m(z), \quad (26)$$

where A is the area of the sample, \mathbf{r} is the position vector in the xy -plane and $\mathbf{R} = (\mathbf{r}, z)$ is a 3D position vector. The matrix element for scattering between states \mathbf{k} and \mathbf{k}' in subbands n and m , respectively, is then given by

$$M(\mathbf{k}, \mathbf{k}')_{nm} = \frac{1}{A} \int e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} d^2\mathbf{r} \int \varphi_m^*(z) H_{qv}(\mathbf{R}) \varphi_n(z) dz \quad (27)$$

where H_{qv} is the interaction potential and the form of the integral with respect to z depends upon the type of the scattering dynamics considered.^{2,34,35} In low-dimensional systems, since the momentum is quantized in one or two directions to form subbands, it is important to note that we now have additional intrasubband and intersubband transitions, which significantly complicates the generation of the scattering tables and choosing the final state after scattering. Below, we give the matrix elements for some of the most important scattering mechanisms that

are present in silicon inversion layers and GaAs/AlGaAs heterostructures.

A. Electron-phonon scattering

Phonon scattering can cause three different types of electronic transitions in the Si inversion layer: transitions between states within a single valley via acoustic phonons (called intravalley acoustic-phonon scattering) and nonpolar optical phonons (called intravalley optical phonon scattering), and transitions between different valleys mediated by high-momentum acoustic or nonpolar optical phonons (called together intervalley scattering).^{36,37,38,39,40,41,42,43} Intravalley acoustic-phonon scattering involves phonons with low energies and is an almost elastic process. The intravalley optical-phonon scattering is induced by optical phonons of low momentum and high energy. Intervalley scattering can be induced by the emission and absorption of high-momentum, high-energy phonons, which can be of either acoustic- or optical-mode variety. Intervalley scattering can therefore be important only for temperatures high enough that an appreciable number of suitable phonons is excited or for hot electrons that can emit high energy phonons.⁴⁴ In order to evaluate the scattering potential that describes the electron-phonon interaction, we need a Hamiltonian that describes the coupled electron-phonon system. The total Hamiltonian of the system is given by^{45,46}

$$\hat{H} = \hat{H}_e + \hat{H}_a + \hat{H}_{ea}, \quad (28)$$

where \hat{H}_e is the electronic part, \hat{H}_a is the atomic part that describes the normal modes of vibration of the solid,

and \hat{H}_{ea} is the electron-ion interaction term of the form

$$\hat{H}_{ea} = \sum_{i,j} V_{ea}(\mathbf{r}_i - \mathbf{R}_j). \quad (29)$$

In general, each ion is at a position $\mathbf{R}_j = \mathbf{R}_j^{(0)} + \mathbf{u}_j$, which is a sum of the equilibrium position $\mathbf{R}_j^{(0)}$ and the displacement \mathbf{u}_j . Under the assumption of small displacements, one can expand V_{ea} in a Taylor series

$$V_{ea}(\mathbf{r} - \mathbf{R}_j) = V_{ea}(\mathbf{r} - \mathbf{R}_j^{(0)}) - \mathbf{u}_j \cdot \nabla_r V_{ea}(\mathbf{r} - \mathbf{R}_j^{(0)}) + o(Q^2). \quad (30)$$

The zero-order term is the potential function for the electrons when the atoms are in their equilibrium positions, which forms a periodic potential in the crystal. The solution of the Hamiltonian for electron motion in this periodic potential gives the Bloch states of the solid. Since the first order term is much smaller than the zero-order term, the electron-phonon interaction can be treated perturbatively. Therefore, the lowest order term for the electron-phonon interaction is of the form

$$H_{e-ph}(\mathbf{r}) = - \sum_j \mathbf{u}_j \cdot \nabla_r V_{ea}(\mathbf{r} - \mathbf{R}_j^{(0)}). \quad (31)$$

It is obvious that this interaction Hamiltonian does not act on the spin variables in this approximation. The Fourier transform of V_{ea} can be written as

$$V_{ea}(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{q}} V_{ea}(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}}, \quad (32)$$

where N is the number of primitive cells, and wavevector \mathbf{q} spans the whole \mathbf{q} -space. Ionic displacement may be decomposed into normal-mode representation and it is customary to write

$$\mathbf{u}_j = i \sum_{\mathbf{k}} \left(\frac{\hbar}{2MN\omega_{\mathbf{k}\lambda}} \right)^{1/2} \times \left(\hat{a}_{\mathbf{k}\lambda} \mathbf{e}_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{R}_j^{(0)}} + \hat{a}_{\mathbf{k}\lambda}^\dagger \mathbf{e}_{\mathbf{k}}^* e^{-i\mathbf{k} \cdot \mathbf{R}_j^{(0)}} \right), \quad (33)$$

where $\mathbf{e}_{\mathbf{k}\lambda}$ is the unit polarization vector that obeys the standard orthonormality and completeness relations, $\omega_{\mathbf{k}\lambda}$ is the phonon frequency of phonon branch λ for wavevector \mathbf{k} (running over the whole Brillouin zone), $\hat{a}_{\mathbf{k}\lambda}$ ($\hat{a}_{\mathbf{k}\lambda}^\dagger$) are the phonon annihilation (creation) operators. In acoustic waves, \mathbf{u}_j refers to the relative displacement of the unit cell as a whole with respect to adjacent unit cells; in optical waves it refers to the relative displacement of the basis atoms within the unit cell. Thus

$$H_{e-ph}(\mathbf{r}) = \sum_{\mathbf{q}, \mathbf{G}} e^{i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} V_{ea}(\mathbf{q} + \mathbf{G}) \times (\mathbf{q} + \mathbf{G}) \cdot \mathbf{e}_{\mathbf{q}\lambda} \left(\frac{\hbar}{2\rho V \omega_{\mathbf{q}\lambda}} \right) (\hat{a}_{\mathbf{q}\lambda} + \hat{a}_{\mathbf{q}\lambda}^\dagger), \quad (34)$$

where $MN = \rho V$, and ρ is the density of the solid. The summation over \mathbf{G} represents summation over all reciprocal lattice vectors of the solid. If one defines a function

$$M_{\mathbf{q}\lambda} = \left(\frac{\hbar}{2\rho V \omega_{\mathbf{q}\lambda}} \right) \sum_{\mathbf{G}} e^{i\mathbf{G} \cdot \mathbf{r}} (\mathbf{q} + \mathbf{G}) \cdot \mathbf{e}_{\mathbf{q}\lambda} V_{ea}(\mathbf{q} + \mathbf{G}), \quad (35)$$

then the Hamiltonian for the electron-phonon interaction becomes

$$H_{e-ph}(\mathbf{r}) = \sum_{\mathbf{q}} M_{\mathbf{q}\lambda} e^{i\mathbf{q} \cdot \mathbf{r}} (\hat{a}_{\mathbf{q}\lambda} + \hat{a}_{\mathbf{q}\lambda}^\dagger). \quad (36)$$

The exact form of the matrix elements for acoustic and nonpolar-optical phonon scattering (zero- and first-order terms) are given below. Since we will need to make a clear distinction between 3D and 2D vectors, *in what follows we are going to use the following notation: capital bold letters will refer to three-dimensional vectors, whereas small bold letters will be used for two-dimensional wavevectors that lie in the xy -plane.*

1. Deformation potential scattering

In general, the application of mechanical stress alters the band structure by shifting energies, and, where it destroys symmetry, by removing degeneracies. It is usually assumed that the mechanical stress does not change the band curvature, and therefore does not change the effective masses, but introduces a shift in the energy states that are close to the band extremum.^{47,48,49,50} For isotropic elastic continuum, the matrix element for deformation potential scattering (acoustic phonons) can be obtained by taking the long-wavelength limit of Eq. (35).⁵¹ For small values of \mathbf{Q} , the summation over reciprocal lattice vectors can be neglected, except for the term $\mathbf{G} = 0$. The screened electron-ion interaction becomes a constant which is usually denoted as Ξ (it gives the shift of the band edge per unit elastic strain). Under these assumptions, $M_{\mathbf{Q}\lambda}$ simplifies to

$$M_{\mathbf{Q}\lambda} = i\mathbf{Q} \cdot \mathbf{e}_{\mathbf{Q}\lambda} \left(\frac{\hbar \Xi^2}{2\rho V \omega_{\mathbf{Q}\lambda}} \right), \quad (37)$$

where $\omega_{\mathbf{Q}\lambda} = V_{s\lambda} Q$, $v_{s\lambda}$ is the sound velocity, is the phonon frequency. Long wavelength acoustic phonons (LA mode) have $\mathbf{Q} \parallel \mathbf{e}_{\mathbf{Q}\lambda}$ which makes the matrix element non-zero. TA phonons have $\mathbf{Q} \perp \mathbf{e}_{\mathbf{Q}\lambda}$ which makes the matrix element vanish. Therefore, the deformation potential mainly couples electrons to LA phonons.

For anisotropic elastic continuum such as silicon, the deformation potential constant Ξ becomes a tensor. The anisotropy of the intravalley deformation potential in the ellipsoidal valleys in silicon has been extensively studied by Herring and Vogt.⁵⁰ Expanding the electron-phonon matrix element over spherical harmonics and retaining only the leading terms, they have expressed the

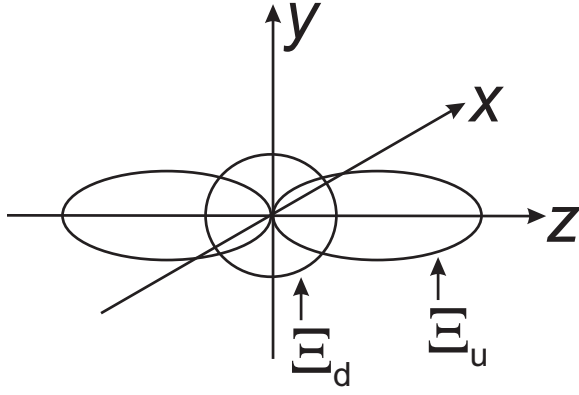


FIG. 11. Angular dependence of the deformation potential for longitudinal modes.

anisotropy of the interaction in terms of the angle $\theta_{\mathbf{Q}}$ between wavevector \mathbf{Q} of the emitted (absorbed) phonon and the longitudinal axis of the valley. They have shown that the matrix element is proportional to Q via the deformation potential $\Delta_{\lambda}(\theta_{\mathbf{Q}})$ ($\lambda = \text{LA}$ or TA) given by^{52,53,54}

$$\Delta_{\text{LA}}(\theta_{\mathbf{Q}}) \approx \Xi_d + \Xi_u \cos^2(\theta_{\mathbf{Q}}), \quad (38)$$

and

$$\Delta_{\text{TA}}(\theta_{\mathbf{Q}}) \approx \Xi_u \cos(\theta_{\mathbf{Q}}) \sin(\theta_{\mathbf{Q}}). \quad (39)$$

Equation (39) accounts for the contribution of both TA branches. Therefore, the acoustic mode scattering is characterized by two constants: Ξ_u (uniaxial shear potential) and Ξ_d (dilatation potential) that is believed to have values of approximately 9.0 eV and -11.7 eV, respectively. In bulk silicon, this anisotropy is usually ignored by using an effective deformation potential constant $\Xi_{\text{LA}}^{\text{eff}}$ for the interaction with longitudinal modes, and ignoring the role of the lower-energy TA modes. This approximation can be justified due to the following reasons: The acoustic modes are most effective at low energy. In this regime and in the usual elastic and equipartition approximation (described later), due to the linear dependence on Q , scattering of electrons at some energy \mathcal{E} samples almost uniformly the constant energy ellipsoid, so that one can take the average values of Δ_{λ} over the ellipsoid. Since there is nothing to fix the energy scale in the problem, this averaging procedure is independent of the electron energy. Moving to the two-dimensional situation, one cannot follow a parallel path to arrive at an isotropic, energy independent effective deformation potential, which complicates the treatment of this scattering process.

Since the wavefunctions of the initial and final states are usually expressed as a product of a one-electron Bloch wavefunction and a harmonic oscillator wavefunction, after the averaging over the phonon states is performed,

the terms inside the brackets of Eq. (36) that represent phonon absorption (term $\hat{a}_{\mathbf{Q}\lambda}$) and phonon emission (term $\hat{a}_{\mathbf{Q}\lambda}^{\dagger}$) processes reduce to $\sqrt{N_{\mathbf{Q}\lambda}}$ and $\sqrt{N_{\mathbf{Q}\lambda} + 1}$, respectively. In thermal equilibrium with a lattice at temperature T , the phonon occupation number $N_{\mathbf{Q}\lambda}$ is given by the Bose-Einstein statistics

$$N_{\mathbf{Q}\lambda} = \frac{1}{e^{\hbar\omega_{\mathbf{Q}\lambda}/k_B T} - 1}, \quad (40)$$

where k_B is the Boltzmann constant. At high enough temperatures, the acoustic phonon energies are much smaller than the thermal energy of electrons. Therefore, one can expand the exponent in the denominator of Eq. (40) into a series and, in the equipartition approximation, appropriate at high temperatures, we have

$$N_{\mathbf{Q}\lambda} + 1 \approx N_{\mathbf{Q}\lambda} \approx \frac{k_B T}{\hbar\omega_{\mathbf{Q}\lambda}} \gg 1. \quad (41)$$

Incorporating these terms as well as the exponential term $e^{iQ_z z}$ into the definition of $M_{\mathbf{Q}\lambda}$, after a straightforward calculation one finds that the matrix element squared for scattering between subbands n and m due to acoustic phonons (for both absorption and emission processes together), and after the averaging over Q_z is performed, reduces to

$$|\langle n | U_{\lambda}^{ac}(\mathbf{q}) | m \rangle|^2 = \frac{k_B T}{\rho V v_{s\lambda}^2} [\Delta_{\lambda, nm}^{\text{eff}}]^2 F_{nm}, \quad (42)$$

where

$$F_{nm} = \int_0^{\infty} dz \psi_n^2(z) \psi_m^2(z). \quad (43)$$

The effective deformation potential constant is calculated from

$$[\Delta_{\lambda, nm}^{\text{eff}}]^2 = \frac{1}{F_{nm}} \int_0^{\infty} dq_z \Delta_{\lambda}^2(\theta_{\mathbf{Q}}) |\mathbf{F}_{nm}(Q_z)|^2, \quad (44)$$

where

$$\mathbf{F}_{nm}(Q_z) = \int_0^{\infty} dz \psi_n(z) \psi_m(z) e^{iQ_z z}. \quad (45)$$

The form-factor $\mathbf{F}_{nm}(q_z)$ introduces an energy scale in the problem by fixing the fuzzy component, the wavevector z . This result is an expected one and follows immediately from the uncertainty principle $\Delta z \Delta p_z \geq \hbar/2$. Since the electrons are frozen into their wavefunctions, and cannot oscillate in the quantized direction, the uncertainty in the particle's location along the z -axis has been reduced. Therefore, there must be a corresponding increase in the uncertainty in the particle's z -directed momentum.³⁴

2. Nonpolar optical phonon scattering

The scattering of electrons by zone-center optical and intervalley phonons in semiconductor crystals has been treated rather extensively by Ferry.^{44,55,56} The nonpolar optical interaction is important for intrasubband scattering as well as for scattering of electrons (and holes) between different minima of the conduction (or valence) band. The latter interaction is important for scattering of carriers in semiconductors with many-valley band structure, such as Si and Ge, and in the Gunn effect, where scattering occurs between different sets of equivalent minima. Harrison⁴⁹ pointed out that the nonpolar optical matrix element may be either of zero or higher order in the phonon wavevector. In subsequent treatments of electron transport in which the nonpolar interaction is important, only the zero-order term was considered, generally owing to the impression that the higher order terms are much smaller. Although this is usually the case, there arise many cases in which the zero-order term is forbidden by the symmetry of the state involved. In these cases, the first order term becomes the leading term, and can become significant in many instances. For example, the first-order intervalley scattering plays an important role in hot-electron transport in the n -type inversion layer in Si. Ignoring this scattering process means that there will be no saturation of the drift velocity at high electric fields, because the zero-order intervalley scattering rate is weakly dependent on the electron energy of high-energy electrons, while the first-order intervalley scattering rate increases as the electron energy increases. The matrix element for nonpolar optical phonon scattering is generally found from a deformable ion model explained in the introduction part of this section. If one thinks of an optical phonon as occurring at finite \mathbf{G} , then the \mathbf{Q} dependence is unimportant, so that the entire matrix element becomes constant and we have

$$M_{\mathbf{Q}\lambda} = M_{0\lambda} = \left(\frac{\hbar D_\lambda^2}{2\rho V \omega_{0\lambda}} \right)^{1/2}, \quad (46)$$

where D_λ is the deformation field (usually given in eV/cm) and $\omega_{0\lambda}$ is the frequency of the relevant phonon mode which is usually taken to be independent of the phonon wave-vector for optical and intervalley processes. Fourier transforming back to real space, a constant in Q -space produces a delta-function in real space. Therefore, this zero-order term represents a short-ranged interaction. A local dilatation or compression of the lattice produces a local fluctuation in the energy of the electron or hole. Incorporating the exponential term $\exp(iQ_z z)$ into the definition of $M_{0\lambda}$, after averaging over Q_z we find

$$\left| \langle n | U_\lambda^{op(0)} | m \rangle \right|^2 = \frac{\hbar D_\lambda^2}{2\rho V \omega_{0\lambda}} F_{nm}, \quad (47)$$

| | $\alpha = 1$ | $\alpha = 2$ |
|--------------|------------------------|------------------------|
| $\alpha = 1$ | $\nu_1 = 1; \nu_2 = 0$ | $\nu_1 = 0; \nu_2 = 4$ |
| $\alpha = 2$ | $\nu_1 = 0; \nu_2 = 2$ | $\nu_1 = 1; \nu_2 = 2$ |

TABLE I. Degeneracy factors for transition between unprimed and primed subbands, for both g- and f-phonons.

for the squared matrix element for scattering between subbands n and m that belong to the α and β valleys, respectively. When the zero-order matrix element for the optical or intervalley interaction vanishes, D_λ is identically zero. In this case, one has to consider the first-order term of the interaction whose matrix element is

$$M_{\mathbf{Q}\lambda} = i\mathbf{Q} \cdot \mathbf{e}_{\mathbf{Q}\lambda} \left(\frac{\hbar D_\lambda^2}{2\rho V \omega_{0\lambda}} \right)^{1/2}. \quad (48)$$

In this context, a first order process means a process similar to acoustic phonon scattering. Following the previously explained procedure, we find that the matrix element squared for scattering between subbands n (α -valley) and m (β -valley) is given by

$$\left| \langle n | U_\lambda^{op(1)} | m \rangle \right|^2 = \frac{\hbar D_{1\lambda}^2}{2\rho V \omega_{0\lambda}} (q^2 F_{nm} + c_{nm}), \quad (49)$$

where

$$c_{nm} = \int_0^\infty dz \left\{ \frac{d}{dz} [\psi_n(z) \psi_m(z)] \right\}^2. \quad (50)$$

The constant term c_{nm} is a small correction term.³⁵

In the scattering among equivalent valleys, there are two types of phonons that might be involved in the process (see Fig. 12). The first type, the so-called g-phonon couples the two valleys along opposite ends of the same axis, i.e. $[100]$ to $[\bar{1}00]$. This is an umklapp process and has a net phonon wavevector $0.3\pi/a$. The f-phonons couple a $\langle 100 \rangle$ valley with $\langle 010 \rangle$, $\langle 001 \rangle$, etc. The reciprocal lattice vector involved in the g-process is \mathbf{G}_{100} and that for an f-process is \mathbf{G}_{111} . Degeneracy factors (g_r) for transition between unprimed ($\alpha = 1$) and primed ($\alpha = 2$) set of subbands, for both g- ($r = 1$) and f-phonons ($r = 2$) are summarized in Table I.

Within a three-subband approximation (subband \mathcal{E}_0 in a $[100]$ valley and \mathcal{E}_1 in $[\bar{1}00]$, and a generic subband \mathcal{E}'_0 in one of the other four valleys – $\langle 010 \rangle$ $\langle 001 \rangle$). Scattering between the \mathcal{E}_0 and \mathcal{E}_1 subbands in the two valleys along the same axis involves only g-type phonons. Scattering between these two minima is usually treated by using a high-energy phonon of 750 K activation temperature (treated as zero-order interaction) and 134-K phonon treated via first order interaction. Scattering between \mathcal{E}_0 or \mathcal{E}_1 and the four \mathcal{E}'_0 subbands involves f-phonons with activation temperatures of 630 K and 230

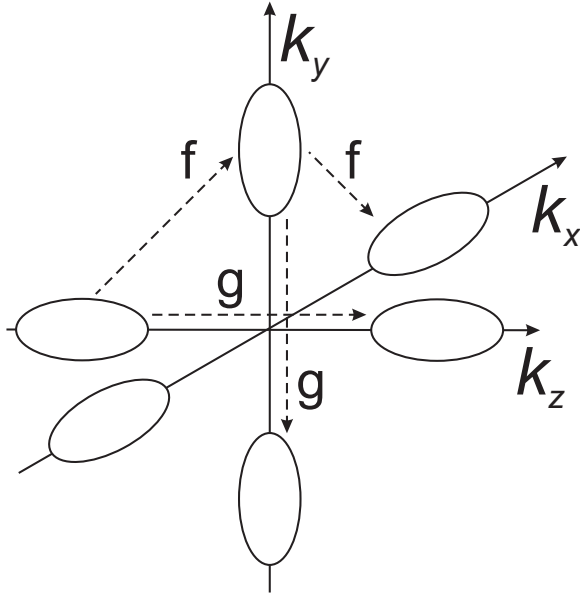


FIG. 12. Diagrammatic representation of intervalley transitions due to g- and f-phonons.

K, treated via zero-order and first-order interaction, respectively. Scattering among subbands \mathcal{E}'_0 involves both g- and f-phonons with activation temperatures of 630 K (zero-order interaction) and 190 K (first-order interaction). All of the high energy phonons are assumed to be coupled with a value of $D_\lambda = 9 \times 10^8$ eV/cm and all of the first-order phonons are assumed to be coupled with $D_{1\lambda} = 5.6$ eV. (This value is consistent with the results given in Ref. 57). The first Born approximation result for the total electron-bulk phonon scattering rate for *p*-type silicon with $N_a = 10^{15}$ cm⁻³, $N_a = 10^{12}$ cm⁻², and T=300 K, with (thick line) and without (thin line) the inclusion of the correction term for the first order process, and for the lowest subband of the unprimed ladder of subbands, is given in Figure 13. We see that, throughout the whole energy range, there is an increase of approximately 10% of the total electron-bulk-phonon scattering rate due to the correction term introduced previously that could lead to mobility reduction. The same trend was also observed for the higher lying subbands.

3. Polar optical phonon scattering

Polar optical phonon (POP) scattering is a very strong scattering mechanism in polar semiconductors, such as GaAs. (It is absent in nonpolar materials, such as Si.) Since our focus here is on Si-based structures, we will not discuss POP scattering in detail, just note that the strength of this mechanism lies in its electrostatic nature. Namely, an optical phonon in a polar binary semiconductor such as GaAs displaces the two atoms in the unit-cell basis with respect to one another, and lead to a modification of the dipole moment associated with

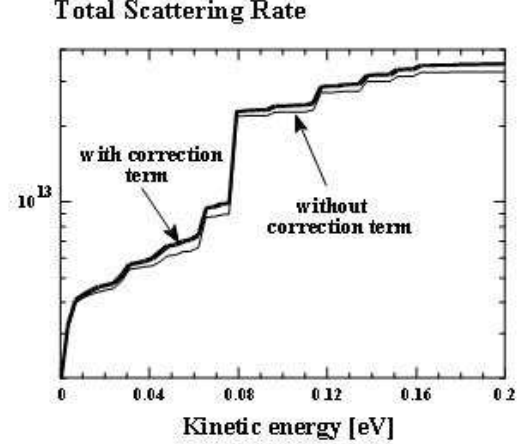


FIG. 13. Total electron-bulk phonon scattering rate for the electrons in the lowest unprimed subband calculated within the first Born approximation.

the unit cell. Overall, the displacement field associated with the propagation of an optical phonon gives rise to an electric field, which is what electrons scatter from. Detailed derivations of the POP scattering rates can be found in many texts.^{2,34} Here we will just give the matrix element squared for POP scattering:

$$|\langle n | U^{pop} | m \rangle|^2 = \frac{\hbar e^2 \omega_{LO}}{2V(q_{||}^2 + q_z^2)} (N_0 + \frac{1}{2} \mp \frac{1}{2}) \quad (51)$$

$$\times \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon(0)} \right) |F_{nm}(q_z)|^2 \delta(k_{||} - k'_{||} \pm q_{||})$$

where $F_{nm}(q_z)$ is given by Eq. (45), ω_{LO} is the longitudinal polar optical phonon frequency and ϵ_∞ and $\epsilon(0)$ are the high frequency and the low frequency dielectric constants, respectively. The scattering rate is then found via integration over the final states and the final expression can be found in Refs. 2 and 34.

B. Scattering by Electrostatic Interactions

1. Coulomb Scattering

Scattering associated with charged Coulomb centers near the plane of the 2D electron gas in MOS devices can be separated into contributions from the depletion layer, the interface charge and the oxide charge. An extensive discussion of the role of multiple-scattering contributions to the electron mobility of a doped semiconductor and the apparent difficulties with the impurity averaging is presented in the papers by Moore,⁵⁸ and Kohn and Luttinger.^{59,60} The expressions for the potential due to a single charge located in the region of interest, in which the image term is properly included, are given by Stern and Howard.⁶¹ Here, we will discuss it in a way that is

suitable for a many-subband treatment. Using the usual method of images,⁶² one easily finds that, in the presence of dielectric medium, the potential due to a charged center located at $\mathbf{R}_i = (\mathbf{r}_i, z_i)$ equals

$$U_i(\mathbf{r}, z) = \frac{e^2}{4\pi\kappa} \frac{1}{\sqrt{(\mathbf{r} - \mathbf{r}_i)^2 + (z - z_i)^2}} \quad (52)$$

for $z_i < 0$, and

$$U_i(\mathbf{r}, z) = \frac{e^2}{4\pi\kappa} \left[\frac{1}{2} \left(1 + \frac{\epsilon_{ox}}{\epsilon_{sc}} \right) \frac{1}{\sqrt{(\mathbf{r} - \mathbf{r}_i)^2 + (z - z_i)^2}} + \frac{1}{2} \left(1 - \frac{\epsilon_{ox}}{\epsilon_{sc}} \right) \frac{1}{\sqrt{(\mathbf{r} - \mathbf{r}_i)^2 + (z + z_i)^2}} \right] \quad (53)$$

for $z_i > 0$. In Eqs. (52) and (53), $\kappa = 0.5(\epsilon_{sc} + \epsilon_{ox})$ is the average dielectric constant at the interface. The depletion charge scattering occurs due to the ionized charges in the depletion layer. Using Eq. (53), we find that the matrix element squared for scattering between subbands n and m due to the depletion charge is equal to

$$|\langle n | U^{\text{depl}}(q) | m \rangle|^2 = |U_{nm}^{\text{depl}}(\mathbf{q})|^2 \quad (54)$$

$$= N_{\text{depl}} \left(\frac{e^2}{2\kappa q} \right)^2 A_{nm}^2(q) \int_0^\infty dz_i O_{nm}^2(q, z_i),$$

where N_{depl} is the depletion charge density, $\mathbf{R}_i = (\mathbf{r}_i, z_i)$ is the location of an arbitrary charge center in the depletion region, and $A_{nm}(q)$ and $O_{nm}(q, z_i)$ are the form factors due to the finite extension of the electron gas in the quantization direction, of the form

$$A_{nm}(q) = \int_0^\infty dz \psi_n(z) e^{-qz} \psi_m(z) \quad (55)$$

and

$$O_{nm}(q, z_i) = 0.5 \left(1 + \frac{\epsilon_{ox}}{\epsilon_{sc}} \right) e^{qz_i} + 0.5 \left(1 - \frac{\epsilon_{ox}}{\epsilon_{sc}} \right) e^{-qz_i} + 0.5 \left(1 + \frac{\epsilon_{ox}}{\epsilon_{sc}} \right) \left[e^{-qz_i} \frac{a_{nm}^{(+)}(q, z_i)}{A_{nm}(q)} - e^{qz_i} \frac{a_{nm}^{(-)}(q, z_i)}{A_{nm}(q)} \right], \quad (56)$$

respectively, where

$$a_{nm}^{(\pm)}(q, z_i) = \int_0^\infty dz \psi_n(z) e^{\pm qz} \psi_m(z). \quad (57)$$

In the above expressions, \mathbf{q} is a wavevector in the plane parallel to the interface (xy -plane in our case).

Near the Si/SiO₂ interface, there are always many Coulomb centers, due to the disorder and defects in the crystalline structure in the neighborhood of the interface. They are associated with the dangling bonds and act as charge-trapping centers which scatter the free carriers through the Coulomb interaction. Using Eq. (52), we find that the matrix element squared for the scattering from a sheet of charge with charge density N_{it} located in the oxide, at distance z_i ($z_i < 0$) from the interface is

$$|\langle n | U^{ij}(\mathbf{q}) | m \rangle|^2 = |U_{nm}^{ij}(\mathbf{q})|^2 \quad (58)$$

$$= N_{\text{it}} \left(\frac{e^2}{2\kappa} \right)^2 \left[\frac{A_{nm}(q)}{q} \right]^2 e^{2qz_i}.$$

For interface-trap scattering $z_i = 0$. By similar arguments, one finds that the matrix element for scattering from the oxide charge, with charge density N_{ox} , is

$$|\langle n | U^{ox}(\mathbf{q}) | m \rangle|^2 = |U_{nm}^{ox}(\mathbf{q})|^2 \quad (59)$$

$$= N_{ox} \left(\frac{e^2}{2\kappa} \right)^2 \left[\frac{A_{nm}(q)}{q} \right]^2 \frac{1 - e^{2qd_{ox}}}{2q},$$

where d_{ox} is the oxide thickness.

For GaAs/AlGaAs heterostructures, $z_i < 0$ and remote Coulomb scattering, in addition to polar optical phonon scattering, dominates the low-field electron mobility. Since there are no charges in the depletion layer in this material system (GaAs is intentionally left undoped), direct Coulomb scattering mechanism can safely be ignored. In addition to the smaller effective mass of electrons in the GaAs system, the absence of direct Coulomb scattering is one of the main reasons for the observation of the very high mobility in GaAs modulation doped heterostructures.

2. Surface-roughness scattering

This scattering mechanism is associated with the interfacial disorder and depends upon the oxidation temperature and ambient as well as post-oxidation anneal and removal of the wafer from the furnace. Early theories of surface roughness were based on the Boltzmann equation in which the surface is incorporated via boundary conditions into the electron distribution function.^{63,64,65} The first quantum-mechanical treatment of the problem was given by Prange and Nee.⁶⁶ Subsequently, the theory followed two different paths. The basic idea of the first approach is to incorporate the variations in the confining potential of the rough surface as a boundary condition on the Hamiltonian of the system. Since there is no simple perturbation theory to treat arbitrary changes in the boundary conditions, the problem of a free-electron Hamiltonian with complicated boundary conditions is then transformed by an appropriate coordinate transformation into a problem with simpler boundary conditions (i.e. into a problem where we have flat surfaces). This coordinate transformation technique has been proposed by Tesanovic *et al.*⁶⁷ and was later used by Trivedi and Ashcroft.⁶⁸ As a consequence of this transformation, the Hamiltonian of the system now has additional terms that play the role of potential interaction terms. These additional terms are treated by perturbative techniques, which are valid when the roughness of the surface is small compared to the thickness of the well.

In the second approach,² the effect of the surface roughness is taken into account through a random local potential term

$$V_0 \Theta[-z + \Delta(\mathbf{r})] - V_0 \Theta(-z) \simeq V_0 \delta(z) \Delta(\mathbf{r}), \quad (60)$$

which is then treated perturbatively (Θ is the step function). The random function $\Delta(\mathbf{r})$ is a measure of the

roughness and is most conveniently expressed in terms of the autocovariance function of $\Delta(\mathbf{r})$. The power spectrum $S(q)$ is the two-dimensional Fourier transform of the autocovariance function of $\Delta(\mathbf{r})$. For the Gaussian correlated roughness that is usually assumed,^{69,70,71,72,73} the power spectrum is given by

$$S_G(q) = \pi\Delta^2\zeta^2 \exp\left(-\frac{q^2\zeta^2}{4}\right). \quad (61)$$

Parameters Δ and ζ characterize the r.m.s. height of the bumps on the surface and the roughness correlation length, respectively. Goodnick *et al.*⁷⁴ have made an extensive analysis of high-resolution transmission electron microscopy (HRTEM) measurements to test the assumption of Gaussian correlation. They found that exponential correlation describes roughness much better than Gaussian correlation irrespective of growth conditions. Roughly speaking, it means that the interface may be regarded as consisting of terraces of a few nanometers in size separated by atomic steps of a few tenths of nanometers, as shown in Fig. 14. This result was later confirmed by Atomic Force Microscope (AFM) measurements.⁷⁵ The power spectrum for the exponential correlation is given by

$$S_E(q) = \frac{\pi\Delta^2\zeta^2}{\left(1 + \frac{q^2\zeta^2}{2}\right)^{3/2}}. \quad (62)$$

A generalization of the result given in Eq. (62) is a self-affine roughness correlation function, which in 2D takes the form

$$S_{SA}(q) = \frac{\pi\Delta^2\zeta^2}{\left(1 + \frac{q^2\zeta^2}{4n}\right)^{n+1}}, \quad (63)$$

where $n > 0$ is an exponent describing the high- q fall-off of the distribution. It reduces to exponential correlation for $n = 0.5$.

For identical roughness parameters, the Gaussian spectrum decays slower for small wavevectors and then falls to zero rapidly for large wavevectors. The exponential model also leads to a rougher interface due to the tails in the spectrum, which allows for short-range fluctuations to be considered as well. For $n < 0.5$ and small values of \mathbf{q} , the power spectrum of the self-affine model decays faster compared to the previous two models, but then falls slowly for large wavevectors. This essentially means that, in this regime, it also allows for short-range fluctuations to be considered. For large exponents, the power spectral density of the self-affine model approaches the one for the Gaussian model. In general, the matrix element for scattering between subbands n and m for this scattering mechanism is of the form

$$|\langle n|U^{sr}(\mathbf{q})|m\rangle|^2 = S(q)\Gamma_{nm}^2(q). \quad (64)$$

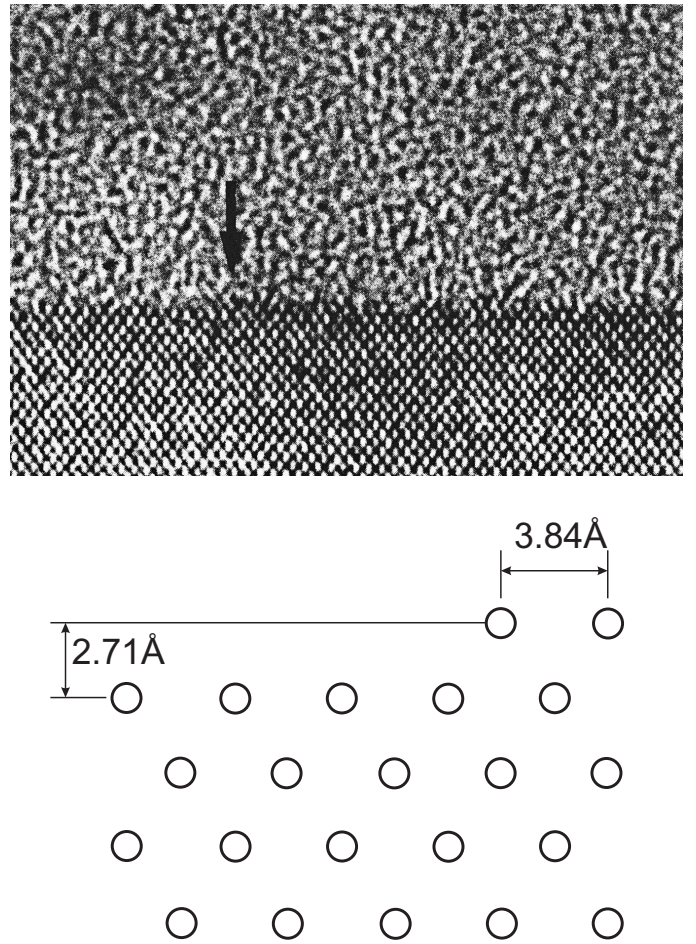


FIG. 14. (Top panel) High-resolution transmission electron micrograph of the interface between Si and SiO₂. The oxide is in the top half of the picture, while the rows of Si atoms can be observed in the bottom half. The image is a lattice plane image lying in the (111) plane, while the interface is a (100) plane. Reprinted with permission from Ref. 74, S. M. Goodnick *et al.*, *Phys. Rev. B* 32, 8171 (1985). © 1985 The American Physical Society. (Bottom panel) Relevant dimensions for the steps occurring at the interface.

For large V_0 , matrix element Γ_{nm} reduces to

$$\begin{aligned} \Gamma_{nm}^{(0)} &= \frac{\hbar^2}{2m_z} \frac{d\psi_n}{dz} \frac{d\psi_m}{dz} \Big|_{z=0} \\ &= \int_0^\infty dz \left\{ \psi_n(z) \frac{\partial V(z)}{\partial z} \psi_m(z) \right. \\ &\quad \left. - \mathcal{E}_m \frac{d\psi_n}{dz} \psi_m(z) + \mathcal{E}_n \psi_n(z) \frac{d\psi_m}{dz} \right\}. \end{aligned} \quad (65)$$

The last expression is the result obtained by Prange and Nee.⁶⁶ Matsumoto and Uemura⁷⁶ calculated that in the electronic quantum limit, $\Gamma_{nm} = eE_{av}$, where $E_{av} \propto \frac{1}{2}N_s + N_{depl}$ (N_s and N_{depl} are the inversion layer and the depletion region sheet charge densities, respectively).

The change in the potential energy of the system due to surface-roughness was corrected by Ando,⁷⁷ by con-

sidering the change in the electron density distribution and the effective dipole moment of the deformed Si-SiO₂ surface. The later scattering rate becomes

$$\Gamma_{nm}(q) = \Gamma_{nm}^{(0)} + \frac{e^2}{\epsilon_{sc} \epsilon_{sc} + \epsilon_{ox}} A_{nm}(q) \quad (66)$$

$$\times \left\{ N_{depl} + N_s - \frac{1}{2} \sum_i N_i A_{ii}(q) \right\},$$

where $\Gamma_{nm}^{(0)}$ is given by Eq. (65). Due to the presence of the dielectric medium, one needs to correct the expression given in Eq. (66) with the contribution of the image term

$$\Gamma_{nm}^{image}(q) = \frac{e^2}{16\pi\epsilon_{sc} \epsilon_{sc} + \epsilon_{ox}} \int_0^\infty dz \psi_n(z) \quad (67)$$

$$\times \left[\frac{K_1(qz)}{qz} - \frac{1}{2} \frac{\epsilon_{sc} - \epsilon_{ox}}{\epsilon_{sc} + \epsilon_{ox}} K_0(qz) \right] \psi_m(z),$$

where K_0 and K_1 are the modified Bessel functions. An additional complication associated with the finite oxide thickness, which may further reduce the mobility via scattering with remote roughness, will be ignored in the present treatment.

3. Electron-Electron Interaction

We consider the Coulomb interaction between an electron with wave vector \mathbf{k} in subband n and a second electron with wave vector \mathbf{k}_2 in subband n_2 . The final states of these two electrons are \mathbf{k}' and n' for the first electron, and \mathbf{k}'_2 and n'_2 for the second electron. The scattering rate for this energy-conserving binary collision process may be written as^{40,78}

$$P^{ee} \left(n, \mathbf{k}; n_2, \mathbf{k}_2 \rightarrow n', \mathbf{k}'; n'_2, \mathbf{k}'_2 \right) \quad (68)$$

$$= \frac{2\pi}{\hbar} |M^{ee}|^2 \delta \left[\mathcal{E}_{n'}(\mathbf{k}') + \mathcal{E}_{n'_2}(\mathbf{k}'_2) - \mathcal{E}_n(\mathbf{k}) - \mathcal{E}_{n_2}(\mathbf{k}_2) \right]$$

where M^{ee} is the matrix element between these two initial and final states. With the inclusion of the exchange effect for indistinguishable particles, the square of the matrix element becomes⁷⁹

$$|M^{ee}|^2 = \frac{1}{4} \left(\left| \langle n', \mathbf{k}'; n'_2, \mathbf{k}'_2 | H^{ee} | n, \mathbf{k}; n_2, \mathbf{k}_2 \rangle \right|^2 \right. \quad (69)$$

$$+ \left| \langle n'_2, \mathbf{k}'_2; n', \mathbf{k}' | H^{ee} | n, \mathbf{k}; n_2, \mathbf{k}_2 \rangle \right|^2$$

$$+ \left| \langle n', \mathbf{k}'; n'_2, \mathbf{k}'_2 | H^{ee} | n, \mathbf{k}; n_2, \mathbf{k}_2 \rangle \right|^2$$

$$\left. - \langle n'_2, \mathbf{k}'_2; n', \mathbf{k}' | H^{ee} | n, \mathbf{k}; n_2, \mathbf{k}_2 \rangle \right|^2 \right),$$

where

$$\left| \langle n', \mathbf{k}'; n'_2, \mathbf{k}'_2 | H^{ee} | n, \mathbf{k}; n_2, \mathbf{k}_2 \rangle \right|^2 \quad (70)$$

$$= |\mathcal{V}(\mathbf{q})|^2 \left| F_{n'n'_2n_2}^{ee}(\mathbf{q}) \right|^2 \delta_{\mathbf{k}'+\mathbf{k}'_2, \mathbf{k}+\mathbf{k}_2}.$$

$\mathcal{V}(\mathbf{q})$ denotes the Fourier transform of the unscreened Coulomb potential and $\mathbf{q} = \mathbf{k}' - \mathbf{k}$ is the exchanged momentum of the first electron with $q = |\mathbf{k}' - \mathbf{k}|$. The form factor is defined by

$$F_{n'n'_2n_2}^{ee}(\mathbf{q}) = \iint dz dz_2 \chi_{n'}(z) \chi_{n'_2}(z_2) \chi_n(z) \chi_{n_2}(z_2) \quad (71)$$

$$\times \exp(-q|z - z_2|).$$

This form factor appears also in the random phase approximation (RPA) of the dielectric function discussed next.

4. Screening of Coulomb, Surface-Roughness and Electron-Electron Scattering

It is known that Coulomb and surface-roughness scattering affect significantly the inversion layer electron mobility, particularly at low and high inversion charge densities. As we said earlier, the scattering potentials for these two dissipative processes are strongly affected by the screening of the mobile charges in the inversion layer. Therefore, any theory that tries to explain the density and temperature dependence of the electron mobility must account for these screening corrections. Since the calculation of the exact dielectric function of homogeneous electron gas is a formidable problem, various approximate solutions for the dielectric function exist in the literature.^{80,81}

Some of these have been very successful, because of their simplicity (Thomas-Fermi method) or high accuracy [the random-phase approximation (RPA)]. The Thomas-Fermi method is basically the semiclassical limit of the Hartree calculation. On the other side, the RPA is an exact Hartree calculation of the charge density in the presence of the self-consistent field of the external charge plus electron gas. More precisely, in the RPA one includes only the long-range Coulomb interaction in the dielectric response, leaving out all exchange-correlation corrections. It leads to the so-called Lindhard dielectric function that is extensively employed in the literature.^{61,82,83,84,85,86,87,88}

IV. TRANSPORT IN QUASI-2D SYSTEMS

The Ensemble Monte Carlo technique has been used for over 30 years now as a numerical method to simulate nonequilibrium transport in semiconductor materials and devices, and has been the subject of numerous books and

reviews.^{4,89,90} In the application to transport problems, a random walk is generated to simulate the stochastic motion of particles subject to collision processes in the medium. This process of random walk generation may be used to evaluate integral equations and is connected to the general random sampling technique used in the evaluation of multi-dimensional integrals.⁹¹ The basic technique is to simulate the free particle motion (referred to as *free flight*) terminated by instantaneous random scattering events. The Monte Carlo algorithm consists of generating random free-flight times for each particle, choosing the type of scattering occurring at the end of the free flight, changing the final energy and momentum of the particle after scattering, and then repeating the procedure for the next free flight. Sampling the particle motion at various times throughout the simulation allows for statistical estimation of physically interesting quantities, such as the single particle distribution function, the average drift velocity in the presence of an applied electric field, the average energy of the particles, etc. By simulating an ensemble of particles representative of the physical system of interest, the non-stationary time-dependent evolution of the electron and hole distributions under the influence of a time-dependent driving force may be simulated.

The particle-based picture, in which the particle motion is decomposed into free flights terminated by instantaneous collisions, is basically the same picture underlying the derivation of the semi-classical Boltzmann transport equation (BTE). In fact, it may be shown that the one-particle distribution function obtained from the random-walk Monte Carlo technique satisfies the BTE for a homogeneous system in the long-time limit.²

A. Mobility Calculation in Silicon Inversion Layers Using 2D Monte Carlo and Including Degeneracy Effects

Multi-particle effects relate to the interaction between particles in the system, which is a nonlinear effect when viewed in the context of the BTE, due to the dependence of such effects on the single particle distribution function itself. Particle-particle interactions are important in Monte Carlo simulation in establishing or relaxing to an equilibrium distribution function characterized by a Maxwell-Boltzmann distribution for non-degenerate situations, or a Fermi-Dirac distribution when proper account for the Pauli exclusion principle is included. Most algorithms developed to deal with such effects essentially linearize the BTE by using the previous value of the distribution function to determine the time evolution of a particle over the successive time-step. Multi-carrier effects may range from simple consideration of the Pauli exclusion principle (which depends on the exact occupancy of states in the system), to single-particle and collective excitations in the system. Inclusion of carrier-carrier interactions in Monte Carlo simulation has been an active

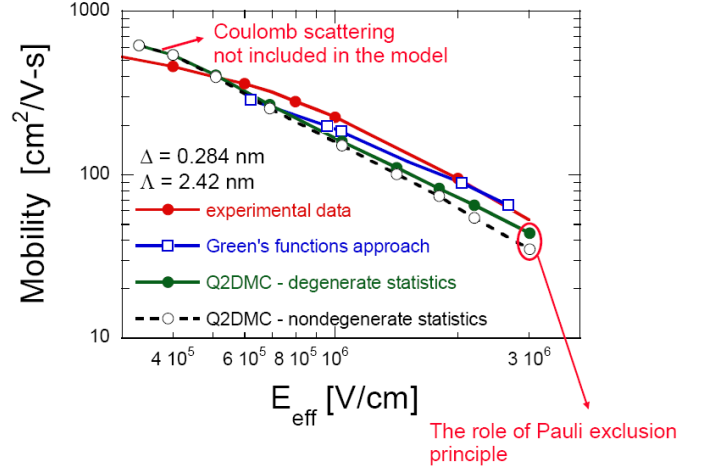


FIG. 15. Mobility vs. sheet electron density for a silicon inversion layer. The wavefunctions of the MOS capacitor structure were calculated using SCHRED³⁰ and fed as input to the 2D Monte Carlo for the solution of the Boltzmann transport equation for confined carriers.

area of research for quite some time and was discussed in conjunction with the description of the scattering mechanisms in the previous section. In that discussion, we only considered binary collisions but higher multi-particle phenomena might be important under high density limit, which are usually referred to as the electron-plasmon interaction and will not be discussed in this text.

Another carrier-carrier effect that is of considerable importance when estimating, for instance, leakage currents in MOSFETs, is impact ionization, which is a pure generation process involving three particles (two electrons and a hole or two holes and an electron). The Pauli exclusion principle requires that the bare scattering rate be modified by a factor $1 - f_m(\mathbf{k})$ in the collision integral of the BTE, where $f_m(\mathbf{k})$ is the one-particle distribution function for the state \mathbf{k} in band (subband) m after scattering. Since the net scattering rate including the Pauli exclusion principle is always less than the bare scattering rate, a self-scattering rejection technique may be used in the Monte Carlo simulation as proposed by Bosi and Jacoboni⁹² for one particle simulation and extended by Lugli and Ferry⁹³ for EMC. In the self-scattering rejection algorithm, an additional random number r is generated (between 0 and 1), and this number is compared to $f_m(\mathbf{k})$, the occupancy of the final state (which is also between 0 and 1 when properly normalized for the numerical \mathbf{k} -space discretization). If r is greater than $f_m(\mathbf{k})$, the scattering is accepted and the particle's momentum and energy are updated accordingly. If this condition is not satisfied, the scattering is rejected, and the process is treated as a self-scattering event with no change of energy or momentum after scattering. Through this algorithm, no scattering to this state can occur if the state is completely full.

Using the rejection technique outlined above, we cal-

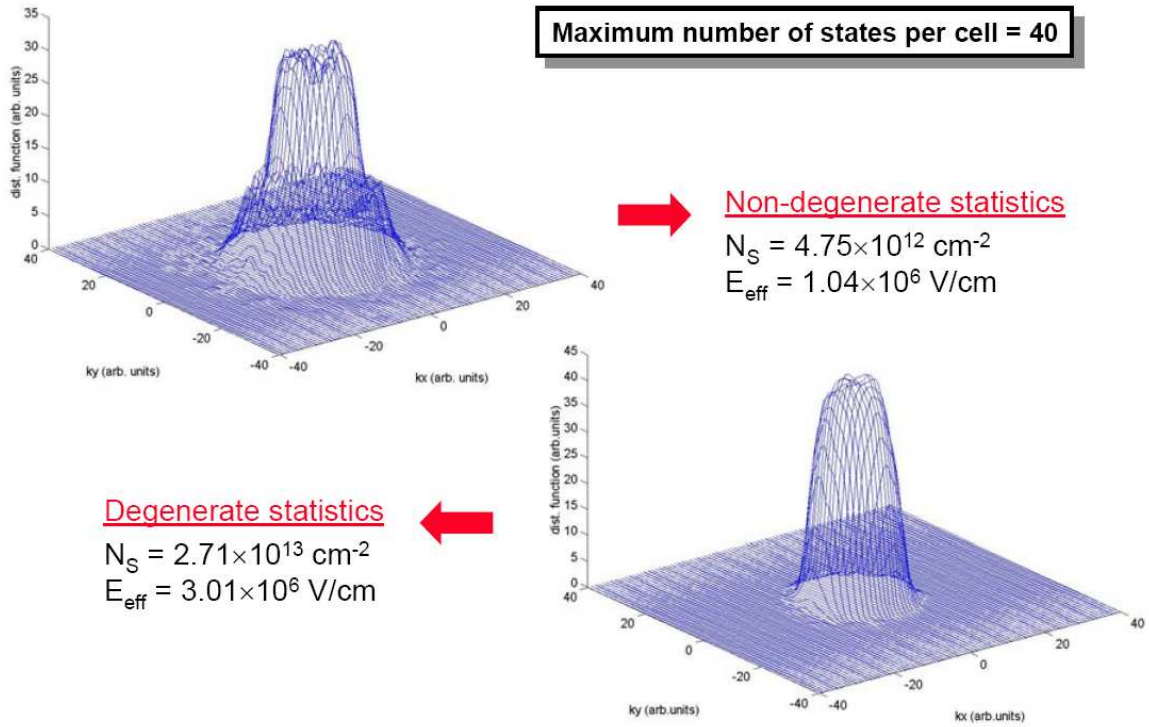


FIG. 16. Distribution functions for a non-degenerate (top) and a degenerate (bottom) quasi-2D electron gas, for given values of the sheet electron density and the effective electric field perpendicular to the semiconductor-oxide interface.

culated the time dependence of the drift velocity along a [100] field direction and used a very low value for the electric field (1kV/cm) so no heating of the carriers was observed. The statistical ensemble consisted of 10,000 particles that were initially distributed amongst different subbands based on the occupancy factors we obtained from SCHRED³⁰ and simulations were performed for 100 ps. The results of the last 50 ps were then time averaged and the mobility was extracted. The variation of the carrier drift mobility vs. effective electric field is given in Figure 15. We see that our simulator clearly reproduces the experimental data and the Green's function results in the mid-to-high effective field region. In these

simulations, we did not include Coulomb scattering and because of that in the low effective field region our simulation results overshoot the experimental values. It is also important to note about a 10% mobility increase in the high sheet electron density region (large effective fields) because of the inclusion of the Pauli exclusion principle which, in turn, leads to closer agreement between the experimental data and the Monte Carlo simulations in the high effective field region. The distribution functions of a non-degenerate and degenerate electron gas are shown in Figure 16. As expected, the degenerate distribution is sharper.

V. QUASI-1D SEMICONDUCTORS (NANOWIRES)

The study of electronic transport in nanowires started almost three decades ago with the seminal work by Sakaki⁹⁴ in 1980. Sakaki showed that the elastic scattering in nanowires (NWs) is suppressed drastically due to a reduction in the final density of states (DOS) for scattering. Electrons in NWs are confined in two transverse directions and are free to move only along the axis

of the wire. In the extreme quantum limit (electrons occupy only the lowest subband), for an elastic scattering process, an electron in the initial state k_i can only be scattered to the final state $k_f = -k_i$ accompanied by a large change in momentum ($2k_i$), since the 1D constant energy surface has only two discrete states. In the case of bulk structures, the constant energy surface is a sphere of radius k_i , so electrons can be scattered to various states including those in the vicinity of k_i . Sakaki considered scattering from remote impurities and showed that the mobility as high as $10^8 \text{ cm}^2/\text{Vs}$ can be achieved in GaAs

NWs at low temperatures because of the substantial reduction of the DOS for scattering in NWs. While the approximations in this work would underestimate the effect of impurities located close to the wire and remote impurity scattering is certainly not the only mechanism limiting the mobility in NWs, the work of Sakaki provided strong impetus for the further study of electronic transport in quasi-1D structures.

In the following year, Arora⁹⁵ calculated the scattering rate due to acoustic phonons and point defects in thin rectangular wire under the relaxation time approximation (RTA) and found them both to increase with decreasing wire cross section. He also showed that the ratio of the conductivity in the wire to the bulk conductivity is proportional to the area of the wire cross section. This result contradicts the previous result of Sakaki, who assumed Coulomb scattering from remote impurities alone in the calculation of the mobility. Lee and Spector⁹⁶ calculated the impurity-limited mobility by accounting for both background and remote impurities using the same approximation as in Sakaki's work. They found the scattering rate from background impurities to be much higher than that from the remote impurities and independent of the wire cross section (probably lost due to the delta-function approximation of wavefunctions along the transverse directions). They also confirmed the acoustic phonon-limited mobility trend in NWs shown by Arora.

Lee and Vassel⁹⁷ considered scattering of electrons from acoustic phonons, impurities (both remote and background), and polar optical phonons and calculated the mobility in wires of various cross sections and different temperatures. For the whole temperature range, they found the mobility in NWs to decrease with decreasing wire cross section. At very low temperatures, where the scattering from impurities dominates transport, mobility in NWs was found to be higher than that in bulk. At room temperature, where phonons dominate electron transport, the mobility in wires of cross section smaller than $12 \times 12 \text{ nm}^2$ was found to be lower than that of the bulk, but for larger cross sections the mobility of the wire was greater than in the bulk.

A. Electronic Bandstructure Modification in Nanowires

All the work discussed above was done on GaAs NWs, with the bulk bandstructure, and assuming unconfined phonons as in the bulk. Sanders *et al.*,⁹⁸ in their work on the electronic transport in free-standing silicon nanowires (SiNWs), indicated yet another important consequence of 2D confinement of electrons in NWs – the modification of the electronic bandstructure because of the change in the dimensionality of the Brillouin zone. The Brillouin zone becomes 1D in NWs since the crystal structure is periodic only along the wire axis. Sanders *et al.* considered SiNWs with axis along $[001]$ and the faces of the wire are $\{110\}$. The primitive cell from which the wire is

constructed is shown in Fig. 17. It contains four silicon atoms (as opposed to only two atoms in a bulk silicon primitive cell); the length of the primitive cell along the axis is a and the transverse dimensions are $a/\sqrt{2}$, where a is the lattice constant ($a = 5.43 \text{ \AA}$). The Brillouin zone in SiNWs is 1D and it extends from $-\pi/a$ to $+\pi/a$, as opposed to $-2\pi/a$ to $+2\pi/a$ in bulk silicon along $[100]$. This is because of the doubling of the length of the unit cell along the wire axis. The conduction band in bulk silicon is indirect and is composed of six equivalent Δ valleys located at $\pm 0.85 \times (2\pi/a) = \pm 1.7\pi/a$ along each of the $\langle 100 \rangle$ directions. In case of the SiNWs with axis along $[001]$, four of the Δ_4 valleys along the transverse directions ($[010]$, $[0\bar{1}0]$, $[100]$, and $[\bar{1}00]$) are projected onto the Γ point in the 1D Brillouin zone and their energies are determined by the effective masses along the $[110]$ and $[\bar{1}10]$ confinement directions. The two Δ_2 valleys along $[001]$ of the bulk Brillouin zone are zone-folded to $\pm 0.3\pi/a$ in SiNWs and become the off- Γ states. The energy bands derived from these are at higher energies than those at the Γ point since the $[001]$ valleys have a lighter effective mass in both confining directions. Thus the SiNW becomes a direct bandgap material.

The work of Sanders *et al.* paved the way for the study of the effects of electronic bandstructure modification on the ballistic transport in silicon nanowire transistors (SWNT) by various groups.^{20,99,100,101,102,103,104} Using the bandstructure obtained from a tight-binding model, Wang *et al.*⁹⁹ studied the validity of using the bulk effective mass in the calculation of electrical properties in SNWTs of different side lengths a . The bulk effective mass was found to overestimate both the threshold voltage (for $a < 3 \text{ nm}$) and the ON-current (for $a < 5 \text{ nm}$). Zheng *et al.*¹⁰⁰ calculated the effective masses, valley splitting between the Δ_4 (at Γ point) and the Δ_2 (off- Γ point) bands, and the bandgap from the full SiNW bandstructure obtained using the tight-binding model. All three were found to increase with decreasing wire cross section. They also found that the single band effective mass equation predicts the bottom of energy bands accurately. Bandstructure in SiNWs is also found to depend on the orientation of the wire axis¹⁰¹ and the species used for surface termination.¹⁰³ For all orientations and surface terminations studied, SiNWs were found to exhibit direct bandgap.

Experimental results on electronic transport in SiNWs were first reported by Cui *et al.* in 2003. The mobility in SiNWs was found to be higher than in bulk silicon. The increase in mobility was attributed to the reduced DOS for scattering as suggested by Sakaki.⁹⁴ In the following year, Koo *et al.*¹⁰⁵ also reported the electron mobility in SiNW FETs to be two times higher than that in bulk MOSFETs, supporting the results of Cui *et al.* In the same year, Kotlyar *et al.*¹⁰⁶ showed that the phonon-limited mobility in SiNWs is much lower than that in the bulk silicon because of the increase in the electron-phonon wavefunction overlap with decreasing cross section. Unlike all previous theoretical studies

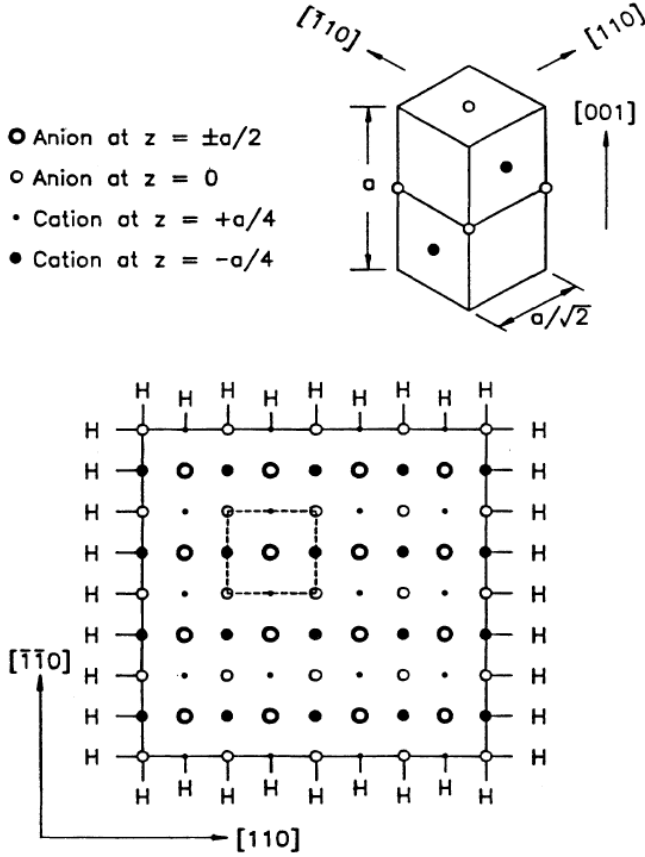


FIG. 17. Crystal structure of an idealized silicon NW. The NW unit cell is shown on the right. Each base unit contains four atoms. The faces are parallel to the four equivalent $\{110\}$ planes and the wire is oriented along $[001]$. Reprinted with permission from Ref. 98, G. D. Sanders, C. J. Stanton, and Y. C. Yang, *Phys. Rev. B* 48, 11067 (1993). © 1993, The American Physical Society.

of transport in NWs, they considered a detailed multi-subband transport and used the electronic wavefunctions and subbands obtained from solving coupled Schrödinger and Poisson equations self-consistently in the calculation of the scattering rates. Wang *et al.* reported that the surface roughness scattering (SRS) becomes less important in the case of ultrasmall SiNWs.¹⁰⁷ Jin *et al.*¹² did a detailed calculation of both phonon-limited and SRS-limited mobility in cylindrical SiNWs and found both to decrease with decreasing wire cross section. A similar mobility reduction in SiNWs has been reported this year by both theoretical^{13,108} and experimental groups.¹⁰⁹

B. Acoustic Phonon Confinement in SiNWs

Another important consequence of 2D spatial confinement is NWs is the modification of the acoustic phonon dispersion in them. In ultrasmall structures such as NWs, the acoustic phonon spectrum is modi-

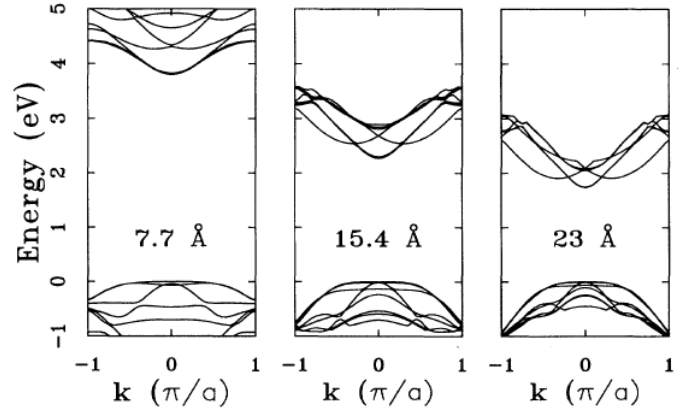


FIG. 18. Bandstructures of square SiNWs of three different widths. The Brillouin zone is 1D with k ranging from $-\pi/a$ to $+\pi/a$. Reprinted with permission from Ref. 98, G. D. Sanders, C. J. Stanton, and Y. C. Yang, *Phys. Rev. B* 48, 11067 (1993). © 1993, The American Physical Society.

fied due to the mismatch of the sound velocities and mass densities between the active layer and the surrounding material,¹¹⁰ in our case Si and SiO₂. This modification in the acoustic phonon spectrum becomes more pronounced as the dimensions of the active layer become smaller than the phonon mean free path, which is around 300 nm in silicon.¹¹¹ Pokatilov *et al.*^{112,113} have shown that the modification in the acoustic phonon dispersion in nanowires can be characterized by the acoustic impedance $\zeta = \rho V_s$, where ρ and V_s are the mass density and sound velocity in the material, respectively. By considering materials with different ζ , Pokatilov *et al.* have shown that the acoustic phonon group velocity in the active layer is reduced when an acoustically soft (smaller ζ) material surrounds an active layer made of acoustically hard (higher ζ) material. Since Si is acoustically harder than SiO₂, the acoustic phonon group velocity in SiNWs with SiO₂ barriers decreases and results in an increased acoustic phonon scattering rate [see Eq. (73)].

The first step in accounting for the acoustic phonon confinement is to calculate the modified acoustic phonon dispersion. Using the adiabatic bond charge model¹¹⁵ (microscopic calculation, accurate but computationally involved), Hepplestone and Srivastava¹¹⁶ have shown the validity of the elastic continuum model (macroscopic calculation, less accurate but easier to implement) for wire dimensions greater than 2.5 nm. Hence, in this work we have used the elastic continuum model to calculate the modified phonon spectrum. Most of the previous studies of acoustic phonon confinement in nanowires have used approximate hybrid modes proposed by Morse¹¹⁷ (valid for wires with thickness much smaller than the width) to calculate the dispersion spectrum. Nishiguchi *et al.*¹¹⁴ calculated the dispersion spectrum using the xyz algorithm¹¹⁸ and found that the Morse formalism is valid only for the lowest phonon subband. Since one acoustic

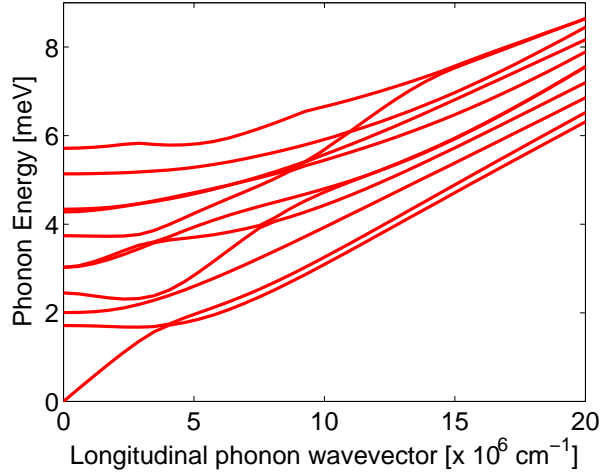


FIG. 19. Confined acoustic phonon dispersion (dilatational mode) calculated using the *xyz* algorithm¹¹⁴ for an 8×8 nm² SiNW. Only the lowest 10 phononic subbands are shown. Dispersion in the first one third of the first Brillouin zone is shown for clarity. Reprinted with permission from Ref. 13, E. B. Ramayya *et al.*, *J. Appl. Phys.* 104, 063711 (2008). © 2008, American Institute of Physics.

phonon subband is surely not enough to accurately describe scattering with electrons, in this work, we have used Nishiguchi *et al.*'s approach to calculate the acoustic phonon dispersion, although it is computationally intensive. The basis functions used to expand the phonon mode displacements in Nishiguchi *et al.*'s approach are powers of Cartesian coordinates, and the number of basis functions required to fully describe the modes depends on the number of modes required. For an 8×8 nm² SiNW, we have found that the lowest 35 phononic subbands are enough to calculate the scattering rate. Also, we found that about 176 basis functions are sufficient to fully describe the displacement of those 35 phononic modes. The number of phononic bands required decreases with the decrease in wire cross section.

Two types of boundary conditions are often used to calculate the acoustic phonon spectrum in nanostructures: a) The free-standing boundary condition (FSBC) assumes that all the surfaces are free, so normal components of the stress tensor vanish at the surfaces, and b) The clamped-surface boundary condition (CSBC) assumes that the surfaces are rigidly fixed, so the displacement of phonon modes is zero at the surfaces. Generally, the CSBC (FSBC) results in higher (lower) phonon group velocity than the bulk case.¹¹⁹ For the wires considered in this work, neither of these boundary conditions holds exactly, since these wires are actually embedded in the SiO₂. Ideally, one needs to solve the elastic continuum equation, taking into account the continuity of displacement and stress at all Si-SiO₂ interfaces, and then apply the boundary conditions at the outer surfaces. But, this is almost numerically impossible for the structure con-

sidered because it would be equivalent to solving the 2D Schrödinger equation in a device with the cross section of about 800×400 nm², three (five if the metal-Si interface is included) interfaces along the depth, and two interfaces along the width. Donetti *et al.*, in their work on a SiO₂-Si-SiO₂ sandwich structure, considered continuity of the displacement and stress at the interfaces to calculate the phonon dispersion.¹¹⁹ They found it to be close to the results from FSBC. Therefore, in this work, we have used the FSBC to calculate the acoustic phonon spectrum of SiNWs. Fig. 19 shows the calculated acoustic phonon dispersion of the lowest 10 dilatational modes for an 8×8 nm² SiNW. Apart from these dilatational modes, depending on the rotational symmetry of the confined acoustic phonon displacement, there are two sets of flexural modes and one set of torsional modes in SiNWs. A detailed description of the symmetry of all these phonon modes can be found in Ref. 114.

VI. SCATTERING IN QUASI-1D ELECTRON SYSTEMS

A. Scattering due to Bulk Acoustic Phonons, Intervalley Phonons, and Surface Roughness

Phonon scattering and the SRS are the most important scattering mechanisms in SiNWs. The SRS was modeled using Ando's model,¹ intervalley scattering was calculated using bulk phonon approximation, and the intravalley acoustic phonons were treated in both the bulk-mode and confined-mode approximations. Since the wire is very lightly doped, the effect of impurity scattering was not included. Nonparabolic band model for silicon, with the nonparabolicity factor $\alpha = 0.5 \text{ eV}^{-1}$, was used in the calculation of scattering rates. A detailed derivation of the 1D scattering rates is given in Appendices of Ref. 13. Here, for brevity, only the final expressions for the scattering rates are given.

For an electron with an initial lateral wavevector k_x and parabolic kinetic energy $\mathcal{E}_{k_x} = \hbar^2 k_x^2 / (2m^*)$ in subband n [with subband energy \mathcal{E}_n and electron wavefunction $\psi_n(y, z)$], scattered to subband m [with subband energy \mathcal{E}_m and electron wavefunction $\psi_m(y, z)$], the final kinetic energy \mathcal{E}_f is given by

$$\mathcal{E}_f = \mathcal{E}_n - \mathcal{E}_m + \frac{\sqrt{1 + 4\alpha\mathcal{E}_{k_x}} - 1}{2\alpha} + \hbar\omega, \quad (72)$$

where $\hbar\omega = 0$ for elastic (bulk intravalley acoustic phonon and surface roughness) scattering, $\hbar\omega = \pm\hbar\omega_0$ for the absorption/emission of an approximately dispersionless intervalley phonon of energy $\hbar\omega_0$, while in the case of confined acoustic phonons (below) the full phonon subband dispersion is incorporated.

The intravalley acoustic phonon scattering rate due

to bulk acoustic phonons is given by

$$\Gamma_{nm}^{ac}(k_x) = \frac{\Xi_{ac}^2 k_B T \sqrt{2m^*}}{\hbar^2 \rho V_s^2} \mathcal{D}_{nm} \frac{(1 + 2\alpha\mathcal{E}_f)}{\sqrt{\mathcal{E}_f(1 + \alpha\mathcal{E}_f)}} \Theta(\mathcal{E}_f), \quad (73)$$

where Ξ_{ac} is the acoustic deformation potential, ρ is the crystal density, V_s is the sound velocity, and Θ is the Heaviside step-function. \mathcal{D}_{nm} represents the overlap integral associated with the electron-phonon interaction (the so-called electron-phonon wavefunction integral,¹⁰⁶) and is given by

$$\mathcal{D}_{nm} = \iint |\psi_n(y, z)|^2 |\psi_m(y, z)|^2 dy dz. \quad (74)$$

For scattering mediated by short wavelength acoustic and optical phonons, the intervalley phonon scattering rate is given by

$$\Gamma_{nm}^{iv}(k_x) = \frac{\Xi_{iv}^2 \sqrt{m^*}}{\sqrt{2} \hbar \rho \omega_0} \left(N_0 + \frac{1}{2} \mp \frac{1}{2} \right) \mathcal{D}_{nm} \times \frac{(1 + 2\alpha\mathcal{E}_f)}{\sqrt{\mathcal{E}_f(1 + \alpha\mathcal{E}_f)}} \Theta(\mathcal{E}_f), \quad (75)$$

where Ξ_{iv} is the intervalley deformation potential, and \mathcal{D}_{nm} is defined in (74). The approximation of dispersionless bulk phonons of energy $\hbar\omega_0$ was adopted to describe an average phonon with wavevector near the edge of the Brillouin zone and $N_0 = [\exp(\hbar\omega_0/k_B T) - 1]^{-1}$ is their average number at temperature T .

Assuming exponentially correlated surface roughness¹²⁰ and incorporating the electron wavefunction deformation due to the interface roughness using Ando's model,¹ the unscreened SRS rate is given by

$$\Gamma_{nm}^{sr}(k_x, \pm) = \frac{2\sqrt{m^*}e^2}{\hbar^2} \frac{\Delta^2 \Lambda}{2 + (q_x^\pm)^2 \Lambda^2} |\mathcal{F}_{nm}|^2 \times \frac{(1 + 2\alpha\mathcal{E}_f)}{\sqrt{\mathcal{E}_f(1 + \alpha\mathcal{E}_f)}} \Theta(\mathcal{E}_f), \quad (76)$$

where Δ and Λ are the r.m.s. height and the correlation length of the fluctuations at the Si-SiO₂ interface, respectively. $q_x^\pm = k_x \pm k'_x$ is the difference between the initial (k_x) and the final (k'_x) electron wavevectors and the top (bottom) sign is for backward (forward) scattering. The SRS overlap integral in Eq. (76) due to the top interface for a silicon body thickness of t_y is given by

$$\begin{aligned} \mathcal{F}_{nm} = & \iint dy dz \left[-\frac{\hbar^2}{et_y m_y} \psi_m(y, z) \frac{\partial^2 \psi_n(y, z)}{\partial y^2} \right. \\ & + \psi_n(y, z) \varepsilon_y(y, z) \left(1 - \frac{y}{t_y} \right) \psi_m(y, z) \\ & \left. + \psi_n(y, z) \left(\frac{\mathcal{E}_m - \mathcal{E}_n}{e} \right) \left(1 - \frac{y}{t_y} \right) \frac{\partial \psi_m(y, z)}{\partial y} \right]. \end{aligned} \quad (77)$$

The SRS overlap integral was derived assuming the interfaces to be uncorrelated. For the bottom interface,

the integration should be performed from the bottom interface to the top interface and the integral for the side interfaces can be obtained by interchanging y and z in Eq. (77). The first term in Eq. (77) is the confinement-induced part of the SRS and it increases with decreasing wire cross section. This term does not depend on the position of the electrons in the channel and hence results in high SRS even at low transverse fields from the gate. The second and third terms in Eq. (77) depend on the average distance of electrons from the interface, so they contribute to the SRS only at high transverse fields from the gate. For wires of cross section smaller than 5×5 nm², major contribution to SRS comes from the confinement-induced term in Eq. (77), and it increases rapidly with decreasing wire cross section.

Scattering rates given by Eqs. (73)–(76) are calculated using the wavefunctions and potential obtained from the self-consistent Poisson-Schrödinger solver. Parameters used for calculating the intervalley scattering were taken from Ref. 121, the acoustic deformation potential was taken from Ref. 20, and $\Delta = 0.3$ nm and $\Lambda = 2.5$ nm were used to characterize the SRS due to each of the four interfaces. The SRS parameters were obtained by fitting the mobility of an 8×30 nm² SiNW in the high transverse field region (where the SRS dominates) with the corresponding mobility observed in ultra-thin SOI of similar thickness.¹²²

B. Scattering due to Confined Acoustic Phonons

The modification of the acoustic phonon dispersion due to confinement, shown in Fig. 19, implies that the linear dispersion and elastic scattering approximation can no longer be used in calculating the scattering rate. The modified scattering rate which takes into account confined acoustic phonon modes is given by

$$\Gamma_{nm}^{ac}(k_x) = \frac{\Xi_{ac}^2}{2WH} \sum_J \sum_{i=1,2} \left(N_{Jq_{x_i}} + \frac{1}{2} \pm \frac{1}{2} \right) \times \frac{1}{\omega_J(q_{x_i}) |\chi_{Jq_{x_i}}^\dagger \mathbf{E} \chi_{Jq_{x_i}}|} \frac{|\mathcal{L}_{nm}(J, q_{x_i})|^2}{|g'(q_{x_i})|}, \quad (78)$$

where q_x is the lateral wavevector of the acoustic phonon, $g(q_x) = \mathcal{E} - \mathcal{E}' \mp \hbar\omega_J(q_x)$, q_{x_1} and q_{x_2} are the two possible roots of $g(q_x) = 0$, and $g'(q_{x_1})$ and $g'(q_{x_2})$ are the derivatives of $g(q_x)$ with respect to q_x evaluated at q_{x_1} and q_{x_2} , respectively. Index J stands for the different acoustic phonon modes and N_{Jq_x} is the number of acoustic phonons of energy $\hbar\omega_J(q_x)$. Overlap integral $\mathcal{L}_{nm}(J, q_x)$ and the total energy of the electron before (\mathcal{E}) and after (\mathcal{E}') scattering are defined in the appendix of Ref. 13. $\chi_{Jq_{x_i}}$ is the eigenvector corresponding to $\omega_J(q_{x_i})$ (for details, see Appendix of Ref. 13).

Fig. 20 shows the electron-acoustic phonon scattering rate calculated using both bulk-mode and confined-

mode approximations. When calculating the electron-bulk acoustic phonon scattering rates, acoustic phonon dispersion is assumed to be linear. This assumption underestimates the electron-acoustic phonon scattering rate with respect to the calculation with confined phonons because the bulk phonon velocity is higher than the velocity of confined phonons and the electron-acoustic phonon scattering rate is inversely proportional to the acoustic phonon group velocity. On average, the confined acoustic phonon scattering rate is about two times the acoustic scattering rate calculated using bulk phonons. Also, each of the bulk-phonon intersubband spikes is split into two groups of spikes corresponding to absorption and emission of different confined phonon modes. In order to account for both intrasubband and intersubband transitions, all four acoustic phonon modes are included in the calculation of the confined acoustic phonon-electron scattering rates. But, it should be noted that the dominant contribution to the scattering rate comes from the dilatational modes (for all other modes the overlap integral $\mathcal{L}_{nm}(J, q_x)$ in Eq.(78) vanishes for intrasubband scattering due to symmetry).

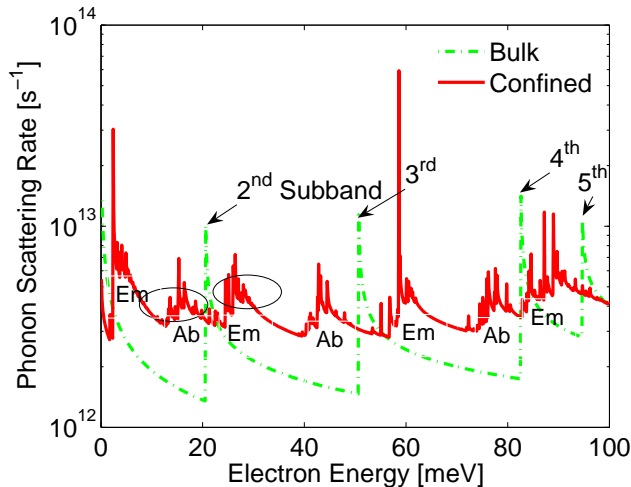


FIG. 20. Electron-acoustic phonon scattering rate of an $8 \times 8 \text{ nm}^2$ SiNW at the channel sheet density of $N_s = 8.1 \times 10^{11} \text{ cm}^{-2}$, calculated assuming the bulk and confined phonon approximations. The electron-bulk acoustic phonon intersubband spikes are at around 20 meV, 52 meV, 85 meV, and 95 meV and they correspond to the electron scattering from the lowest subband to the 2nd, 3rd, 4th, and the 5th lowest subbands. The set of spikes indicated by *Ab* and *Em* correspond to the absorption and emission of phonons from different phonon subbands. Reprinted with permission from Ref. 13, E. B. Ramayya *et al.*, *J. Appl. Phys.* 104, 063711 (2008). © 2008, American Institute of Physics.

VII. TRANSPORT IN QUASI-1D ELECTRON SYSTEMS

In the following, we present recent calculations^{13,123,124} of the electron mobility in rectangular (Sec. VII A) and square (Sec. VII B) SiNWs, in which transport is governed by scattering of electrons with acoustic phonons (confined and bulk), intervalley phonons, and imperfections at the Si-SiO₂ interface. We consider wires much longer than the electron mean-free path under a low lateral electric field. The nanowires are formed on thin silicon-on-insulator (SOI), with the gate oxide, buried oxide, and bottom silicon substrate thicknesses equal to 25 nm, 80 nm, and 700 nm, respectively. All gated wires (rectangular and square) have 200 nm of oxide on the side. The channel is lightly doped to $3 \times 10^{15} \text{ cm}^{-3}$ and the wires are assumed homogeneous and infinitely long. In Sec. VII C, we also briefly discuss transport in highly doped ungated wires used for thermoelectric applications, which have only 2 nm of native oxide all around.

Bulk-silicon effective mass parameters are used in the calculation of the scattering rates. The confined acoustic phonon spectrum is obtained by using the *xyz* algorithm^{114,118} and the unscreened SRS is modeled using modified Ando's model¹ that accounts for the finite thickness of the silicon layer. The simulator used to calculate the electron mobility has two components: the first is a self-consistent 2D Poisson-2D Schrödinger solver and the second is a Monte-Carlo transport kernel.^{13,123,124} The former is used to calculate the electronic states and the self-consistent potential distribution across the wire and the latter simulates transport along the wire axis. The finite barrier at the Si-SiO₂ interface results in the electron wavefunction penetration through the interface and into the oxide. The wavefunction penetration is accounted for by including a few mesh points in the oxide while solving the Schrödinger equation. ARPACK package¹²⁵ was used to solve the 2D Schrödinger equation and successive over-relaxation (SOR) method was used to solve the 2D Poisson equation. The convergence of the coupled Schrödinger-Poisson solver is found to be faster when the Poisson equation is solved using the SOR method than when it is solved using incomplete LU (ILU) method.

A Monte Carlo transport kernel is used to simulate electron transport along the axis of the wire under the influence of the confining potential in the transverse directions and a very small lateral electric field along the channel. The long wire approximation implies that the transport is diffusive (the length exceeds the carrier mean free path), and therefore justifies the use of the Monte Carlo method^{4,126} to simulate electron transport. Electrons are initialized such that their average kinetic energy is $(1/2)k_B T$ (thermal energy for 1D) and are distributed among different subbands obtained from the Poisson-Schrödinger solver in accordance with the equilibrium distribution. Since the electrons are confined in

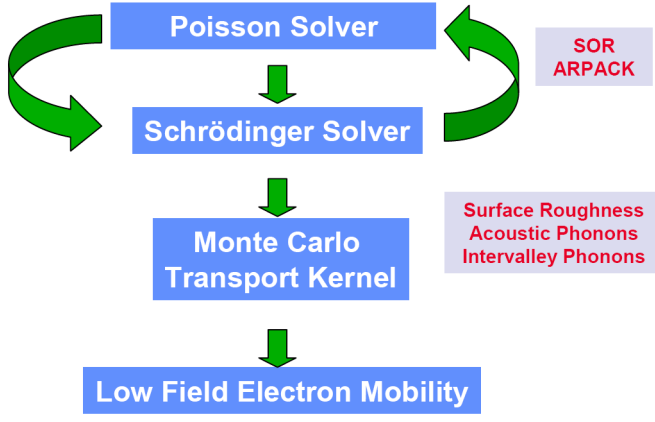


FIG. 21. Flowchart of the simulator developed to calculate the electron mobility in SiNWs.

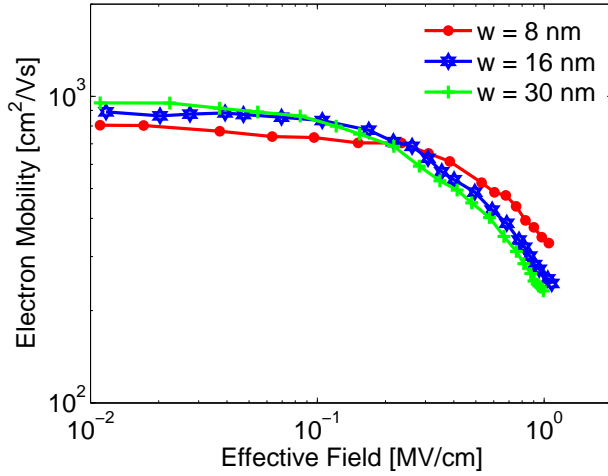


FIG. 22. Variation of the field-dependent mobility with varying SiNW width. The wire thickness is kept constant at 8 nm. Reprinted with permission from Ref. 123, E. B. Ramayya *et al.*, *IEEE Trans. Nanotechnol.* 6, 113 (2007). © 2007, IEEE.

two transverse directions, they are only scattered either forward or backward; consequently, just the carrier momentum along the length of the wire needs to be updated after each scattering event. Mobility is calculated from the ensemble average of the electron velocities.⁴

A. Gated Rectangular Nanowires

The variation of the field-dependent mobility with decreasing channel width was investigated on a series of SiNWs, while keeping the channel thickness at 8 nm and using the bulk phonon approximation. Fig. 22 shows the mobility for SiNWs with the widths of 30 nm, 16 nm and 8 nm. Two important results regarding the mobility behavior in the width range considered can be deduced from Fig. 22: (i) the mobility at high transverse fields,

which is dominated by the SRS, increases with decreasing wire width and (ii) the mobility at low-to-moderate transverse fields, determined by phonon scattering, decreases with decreasing wire width. In the calculation of the SRS overlap integral, initially infinite wire thickness was assumed as in the original Ando's derivation. With $t_s \rightarrow \infty$ assumption the integral becomes

$$\mathcal{F}_{nm} = \iint dy dz [\psi_n(y, z) \varepsilon_y(y, z) \psi_m(y, z) + \psi_n(y, z) \left(\frac{\mathcal{E}_m - \mathcal{E}_n}{e} \right) \frac{\partial \psi_m(y, z)}{\partial y}] \quad (79)$$

Phonon scattering variation with decreasing wire width is determined by the interplay of two opposing factors: (i) reduction of the final density of states for the electrons to scatter to, and (ii) an increase in the electron-phonon wavefunction overlap (74). The former results in an enhanced mobility, while the latter results in mobility degradation. The overlap integral (74) increases with a decrease in the wire width due to an increase in the electron confinement.¹²³ In narrow wires, the increase in the electron-phonon wavefunction overlap dominates over the density-of-states reduction, resulting in a net decrease in the electron mobility at low-to-moderate transverse fields.

The SRS overlap integral given by (77) has three terms. Two are due to deformation of the wavefunction, and the third, dominant term depends on the strength of the field perpendicular to the interface. Since the field normal to the side interfaces is very weak, SRS due to these interfaces is much less efficient than scattering due to the top and bottom ones. The decrease in SRS with decreasing wire width can be understood by following the behavior of the average distance (80) of the carriers from the top interface:

$$\langle y \rangle = \frac{1}{N_l} \sum_{i,v} N_l^{i,v} \iint |\psi_i^v(y, z)|^2 y \, dy \, dz, \quad (80)$$

where N_l is the total line density and $N_l^{i,v}$ is the line density in the i^{th} subband of the v^{th} valley. In Fig. 23, we can see that the carriers are moving away from the top interface as the width of the wire is decreased, and are therefore not strongly influenced by the interface. This behavior is also observed in the case of ultrathin double-gate SOI FETs, and is due to the onset of volume inversion.^{127,128} As the SiNW approaches the volume inversion limit, carriers cease to be confined to the interfaces, but are distributed throughout the silicon volume. Fig. 24 shows the distribution of carriers in a 30 nm wire and an 8 nm wire at the same effective field; we can clearly see that the carriers are confined extremely close to the interface in the 30 wide nm wire, whereas they are distributed throughout the silicon layer in the case of the 8 nm wire.

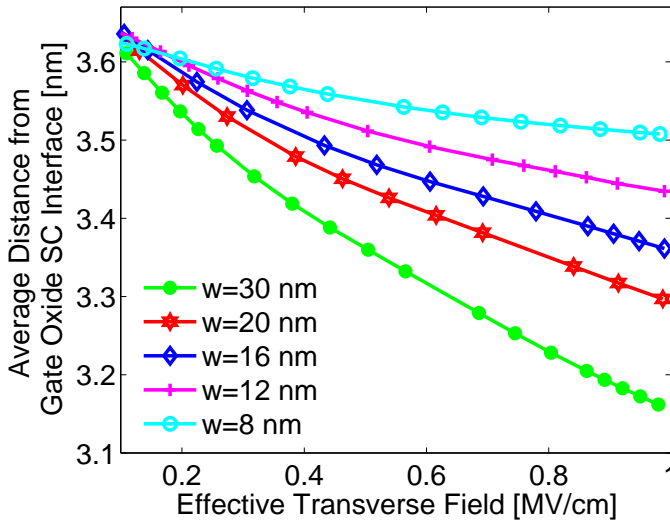


FIG. 23. Variation of the average distance of carriers from the top interface (below the gate) for various wire widths as a function of the effective transverse field. Reprinted with permission from Ref. 123, E. B. Ramayya *et al.*, *IEEE Trans. Nanotechnol.* 6, 113 (2007). © 2007, IEEE.

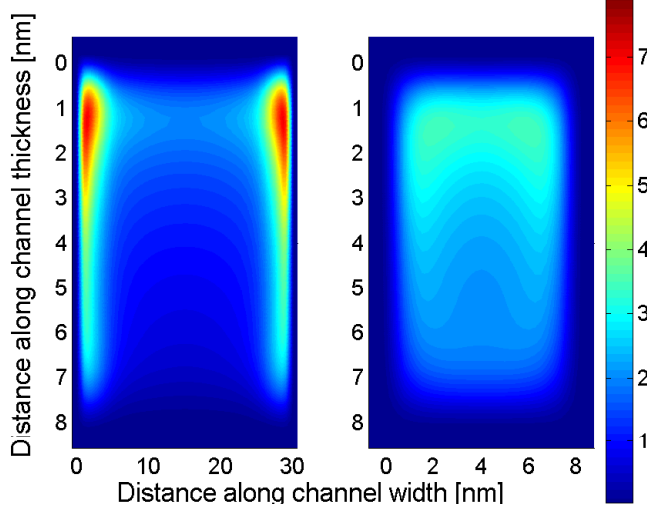


FIG. 24. Electron distribution across the nanowire, for the wire width of 30 nm (left panel) and 8 nm (right panel). In both panels, the transverse field is 1 MV/cm, the wire thickness is 8 nm, and the color scale is in $\times 10^{19} \text{ cm}^{-3}$. Reprinted with permission from Ref. 123, E. B. Ramayya *et al.*, *IEEE Trans. Nanotechnol.* 6, 113 (2007). © 2007, IEEE.

B. Gated Square Nanowires

In this section, we emphasize the importance of acoustic phonon confinement in SiNWs, and vary the cross section of the wire to investigate the effect of increasing spatial confinement on electron mobility. The SRS overlap integral given by Eq. (77) is used to calculate the SRS to account for the finite thickness and width of the wire.

The electron mobility in an $8 \times 8 \text{ nm}^2$ wire, with and

without phonon confinement, is shown in Fig. 25. In the low transverse field regime, the mobility calculated with confined acoustic phonons is about 10 % smaller than that obtained with bulk phonons. This clearly indicates that confined acoustic phonons need to be properly included in the study of electrical transport in SiNWs. The mobility values for $8 \times 8 \text{ nm}^2$ are very close to the experimentally observed mobility in an ultra-thin SOI of similar thickness.¹²² In the rest of the section, acoustic phonons are treated under confined mode approximation.

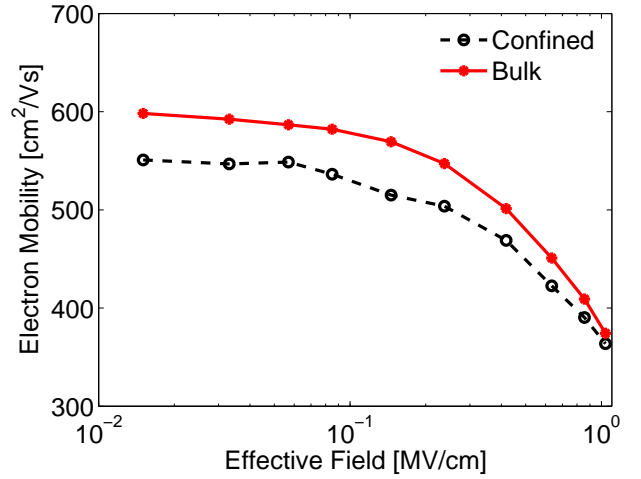


FIG. 25. Variation of the field-dependent mobility for an $8 \times 8 \text{ nm}^2$ SiNW assuming bulk acoustic phonons (solid line) and confined acoustic phonons (dashed line). Reprinted with permission from Ref. 13, E. B. Ramayya *et al.*, *J. Appl. Phys.* 104, 063711 (2008). © 2008, American Institute of Physics.

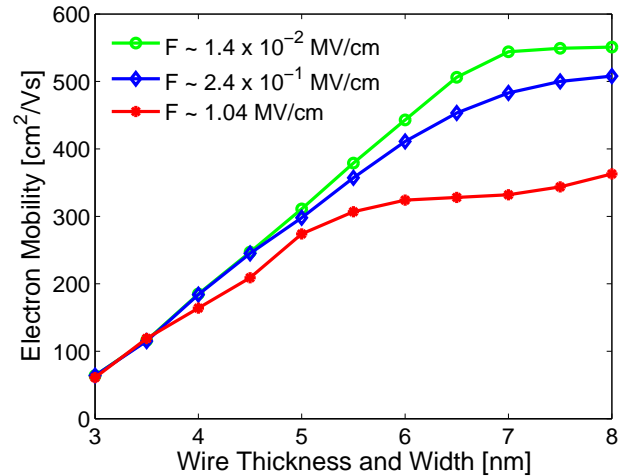


FIG. 26. Variation of the electron mobility with SiNW cross section at three different transverse fields. Reprinted with permission from Ref. 13, E. B. Ramayya *et al.*, *J. Appl. Phys.* 104, 063711 (2008). © 2008, American Institute of Physics.

To determine the cross sectional dependence of elec-

tron mobility in SiNWs and to understand the confinement effects on the spatial and k -space distribution of electrons, the cross section of the wire was varied from $8 \times 8 \text{ nm}^2$ to $3 \times 3 \text{ nm}^2$. The variation of the electron mobility with decreasing wire cross section at a low ($1.4 \times 10^{-2} \text{ MV/cm}$), moderate ($2.4 \times 10^{-1} \text{ MV/cm}$) and high (1.04 MV/cm) transverse field is plotted in Fig. 26.

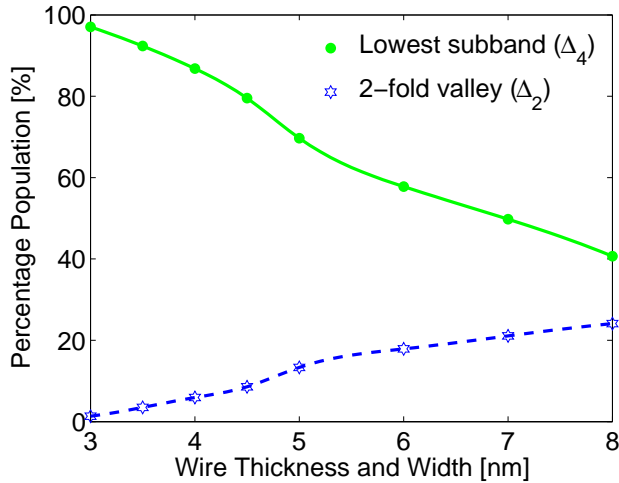


FIG. 27. Variation of the electron population at a low transverse field ($1.4 \times 10^{-2} \text{ MV/cm}$). The solid line shows the total population of the electrons in the lowest subband in each of the four Δ_4 valley pairs and the dashed line shows population of the Δ_2 valley pair, with varying spatial confinement. Reprinted with permission from Ref. 13, E. B. Ramayya *et al.*, *J. Appl. Phys.* 104, 063711 (2008). © 2008, American Institute of Physics.

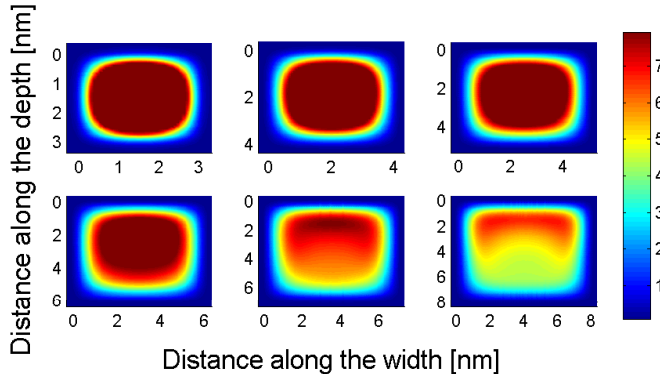


FIG. 28. Electron density across the nanowire at a high transverse field (1.04 MV/cm). When the cross section is reduced from $8 \times 8 \text{ nm}^2$ (bottom right panel) to $3 \times 3 \text{ nm}^2$ (top left), the onset of volume inversion is evident. The color scale is in $5 \times 10^{18} \text{ cm}^{-3}$. Reprinted with permission from Ref. 13, E. B. Ramayya *et al.*, *J. Appl. Phys.* 104, 063711 (2008). © 2008, American Institute of Physics.

1. Subband Modulation

One of the important factors that determine the energy and occupation probability of a subband in each of the Δ_6 valleys (equivalent in bulk silicon) is the effective mass in the direction of confinement. For the SiNWs considered, the confinement is along the y and z directions and the electrons are allowed to move freely in the x direction; consequently, the conductivity effective mass for the valley pairs with minima on the x , y , and z axes are m_l , m_t , and m_t , respectively, while their subband energies are proportional to $1/\sqrt{m_t^2}$, $1/\sqrt{m_t m_l}$, and $1/\sqrt{m_t m_l}$, respectively. Since $m_t < m_l$, the subbands in the valley pair along x are higher in energy than those in the valley pairs along y and z . So the subbands split into those originating from the twofold degenerate Δ_2 (the valley pair along x) and those originating from the fourfold degenerate Δ_4 valleys (the valley pairs along y and z). Upon increasing spatial confinement by decreasing the wire cross section, the subbands in different valleys are pushed higher up in energy, and consequently only a few of the lowest subbands in each of the valley pairs get populated with electrons. Fig. 27 shows the depopulation of the higher Δ_2 valley subbands with increasing spatial confinement: since the lowest subbands in the Δ_4 valleys are lower in energy than those in the Δ_2 valleys, under extreme confinement Δ_2 subbands get completely depopulated, and only the lowest Δ_4 subbands are populated. Splitting of the valley degeneracy and modification of the subband energies in different valley pairs due to spatial confinement, followed by depopulation of the higher subbands, are together termed *subband modulation*.¹²¹ Although subband modulation results in enhanced electron mobility in ultrathin-body SOI MOSFETs,¹²² the rapid increase in SRS for wire cross section below $5 \times 5 \text{ nm}^2$ suppresses the effect of subband modulation in ultra-small SiNWs. In our previous work,¹²⁴ we did observe a small enhancement in mobility for wires of cross section around $4 \times 4 \text{ nm}^2$, but in that study we did not consider the confinement induced term in SRS. This has been shown to be the dominant term in determining the SRS in ultra small SiNWs.¹²

2. Volume Inversion

As the cross section of the SiNW decreases, the channel electrons are distributed throughout the silicon volume as opposed to just within a thin channel at the Si-SiO₂ interface in conventional MOSFETs. The transition from surface inversion to volume inversion occurs gradually and the cross section at which the entire silicon is inverted depends on the electron sheet density.¹²⁹ Fig. 28 shows the variation of the electron density across the wire with varying wire dimensions. When the cross section is decreased from $8 \times 8 \text{ nm}^2$ to about $6 \times 6 \text{ nm}^2$, the onset of volume inversion results in an increase in the average distance of the electrons from the interfaces.

But when the wire cross section is around $6 \times 6 \text{ nm}^2$, the silicon layer is fully inverted, so further decrease in the cross section simply results in a decrease of the average distance of the electrons from the interfaces, thereby resulting in more SRS. Consequently, for wires with the cross section smaller than $6 \times 6 \text{ nm}^2$, volume inversion does not offer an advantage to electronic transport.

3. Mobility Variation with the SiNW Cross Section

Fig. 26 shows the mobility variation when the SiNW cross section is varied from $8 \times 8 \text{ nm}^2$ to $3 \times 3 \text{ nm}^2$. The three curves correspond to three different transverse effective fields. Irrespective of the effective field from the gate, the mobility decreases with increasing spatial confinement mainly due to the monotonic increase in the confinement-induced part of SRS and the intrasubband phonon scattering. At low and moderate effective fields from the gate, the mobility is determined by the scattering from phonons and the confinement-induced part of SRS, whereas at high fields, mobility is limited by SRS (all three terms in the SRS overlap integral, given by Eq.(77) play crucial roles).

At high fields from the gate, when the wire cross section is reduced from $8 \times 8 \text{ nm}^2$ to $5 \times 5 \text{ nm}^2$, the first term in the SRS overlap integral increases, whereas the second and the third terms decrease due to the onset of volume inversion (Fig. 28). Consequently, the mobility shows a very small change for these cross sections. But, when the wire cross section is smaller than $5 \times 5 \text{ nm}^2$, all the terms in the SRS overlap integral increase with decreasing wire cross section resulting in a monotonic decrease in mobility with increasing spatial confinement.

At low and moderate fields from the gate, with decreasing wire cross section: intrasubband phonon scattering increases due to the increase in the electron-phonon overlap integral; intersubband scattering and intervalley phonon scattering decreases due to subband modulation; SRS increases due to the increase in the first term in the SRS overlap integral with increasing confinement. Overall the mobility decreases with decreasing wire cross section. It should be noted that at low transverse fields, the mobility variation for wires of cross section larger than $7 \times 7 \text{ nm}^2$ becomes very small. This is because of the interplay between a simultaneous increase in intersubband scattering (number of occupied subbands increases) and a decrease in intrasubband scattering (electron-phonon overlap integral decreases) with increasing wire cross section. A similar weak dependence of the electron mobility on the wire cross-sectional dimension has been reported by Jin *et al.*¹² for cylindrical SiNW of diameter above 6 nm.

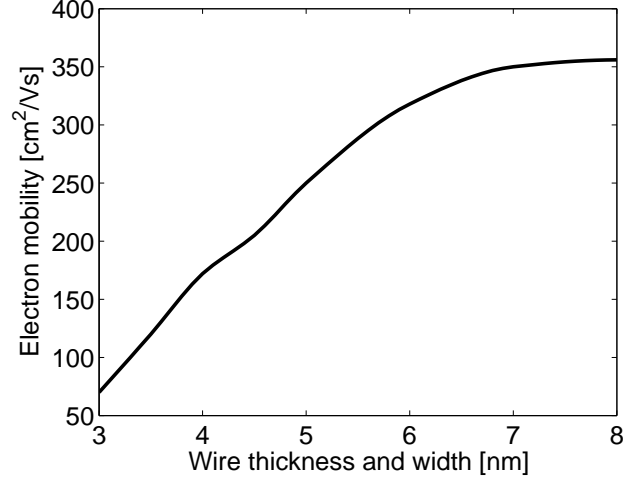


FIG. 29. Variation of the electron mobility with the wire cross section in ungated square SiNWs. The trend in the mobility variation is similar to the trend observed in gated wires at a very low transverse field.

C. Ungated Silicon Nanowires

In contrast to the wires considered so far, SiNWs used for thermoelectric applications are ungated. They are also very highly doped to increase the electrical conductivity. In obtaining the self-consistent wavefunctions and potential in ungated wires, the electric field and the wavefunctions are forced to zero at the air-SiO₂ interfaces. The silicon channel is assumed to be doped to $1.6 \times 10^{19} \text{ cm}^{-3}$ *n*-type with arsenic, and the native oxide thickness is 2 nm all around. Due to the high doping concentration, scattering from ionized impurities is expected to play a crucial role in determining the mobility, so in addition to the SRS and phonon scattering considered before, scattering due to impurities is also included in the mobility calculation.

The electron-impurity scattering rate is given by

$$\Gamma_{nm}^{imp}(\mathbf{k}_x) = \frac{Z^2 e^4 N_a \sqrt{m}}{16 \sqrt{2} \pi^2 \hbar^2 \epsilon_{si}^2} \frac{(1 + 2\alpha \mathcal{E}_f)}{\sqrt{\mathcal{E}_f(1 + \alpha \mathcal{E}_f)}} \times \int d\mathbf{R} I_{nm}^2(q_x^\pm, \mathbf{R}), \quad (81)$$

where $Z = 1$ is the number of electrons donated by the impurity atom, N_a is the doping density, ϵ_{si} is the dielectric constant of silicon, e is the electron charge, m is the electron mass, α is the non-parabolicity factor, \mathcal{E}_f is the final kinetic energy as defined in Eq. (72), $q_x^\pm = k_x \pm k'_x$ is the difference between the initial and final electron wavevectors, and \mathbf{R} is the position of the impurity in the wire cross section. The impurity overlap integral $I_{nm}^2(q_x^\pm, \mathbf{R})$ is defined as

$$I_{nm}(q, \mathbf{R}) = \int \psi_n(y, z) K_0(q, \mathbf{r}, \mathbf{R}) \psi_m(y, z) dy dz, \quad (82)$$

where

$$K_0(q, \mathbf{r}, \mathbf{R}) = \int \frac{e^{iqx} e^{-\frac{\sqrt{(\mathbf{r}-\mathbf{R})^2+x^2}}{L_d}}}{\sqrt{(\mathbf{r}-\mathbf{R})^2+x^2}} dx. \quad (83)$$

Since the doping level is in the degenerate limit, the Pauli exclusion principle is included in the Monte Carlo simulation via the rejection technique described earlier (Sec. IV): an electron is allowed to undergo a given scattering event only if the final state is empty, and if not, the scattering event is treated as self-scattering.¹³⁰ Electron mobility in ungated wires is found to decrease with decreasing wire cross section (Fig. 29), as previously seen in gated wires at low transverse fields (Fig. 26).

VIII. CONCLUDING REMARKS

In this paper, we have reviewed the semiclassical transport description in Q2DEGs and Q1DEGs. Spatial confinement in Q2DEGs and Q1DEGs leads to formation of low-dimensional subbands, and a system of coupled Schrödinger and Poisson equations in the effective mass approximation needs to be solved to obtain the subband energies. Subbands are then populated according to the Boltzmann transport equation, whose state-of-the-art numerical solution is based on the ensemble Monte Carlo technique. Ensemble Monte Carlo remains fairly robust for room-temperature transport description down to deep-submicron scales.

Representative examples – the silicon MOSFET and silicon nanowires with varying cross section – were analyzed in detail. As the MOSFET oxide thickness is scaled to below 10 nm, quantum confinement of inversion charge leads to an appreciable inversion layer capacitance in series with the oxide, so the total gate capacitance is lowered. The influence of the inversion layer capacitance on

the threshold voltage shift and device transconductance has been reliably modeled by self-consistent Schrödinger-Poisson solvers. Furthermore, the field-dependent mobility curve for the silicon MOSFET obtained in experiment has been reproduced very well by using a Schrödinger-Poisson-Monte Carlo transport simulation of the inversion layer Q2DEG. The importance of accounting for the degeneracy in the carrier statistics was clearly demonstrated.

In ultrathin SiNWs, both electrons and acoustic phonons experience 2D confinement. In wires surrounded by SiO₂, an acoustically softer material, the acoustic phonon group velocity is lowered to almost half of its bulk silicon value, and leads to enhanced electron-acoustic phonon scattering rates. The electron mobility calculated while accounting for the modification to the acoustic phonon spectrum due to confinement is about 10% lower than the mobility calculated with bulk acoustic phonons. For very thin wires (below the 5×5 nm² cross section), the mobility decreases monotonically with increasing spatial confinement and becomes virtually independent of the transverse electric field. This occurs primarily due to the increase in the field-independent, confinement-induced part of the SRS, and second due to the increase in intrasubband phonon scattering. In contrast to bulk MOSFETs, in which the SRS plays an important role only for high fields from the gate, electrons in very thin SiNWs are strongly influenced by the roughness regardless of the transverse field. This finding is important for field-effect transistors with multiple gates, as well as for ungated ultrathin wires used for thermoelectric applications or interconnects.

Acknowledgment

IK and EBR acknowledge support from the National Science Foundation, awards ECCS-0547415 and DMR-0520527. DV and SMG acknowledge support from the Arizona Institute for Nano Electronics (AINE).

¹ T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

² D. K. Ferry, *Semiconductors* (MacMillan, New York, 1991).

³ D. K. Ferry and S. M. Goodick, *Transport in nanostructures* (Cambridge, Cambridge, UK, 1997).

⁴ C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).

⁵ M. Nedjalkov, D. Vasileska, D. K. Ferry, C. Jacoboni, C. Ringhofer, I. Dimov, and V. Palankovski, Phys. Rev. B **74**, 035311 (2006).

⁶ R. Lake and S. Datta, Phys. Rev. B **45**, 6670 (1992).

⁷ R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).

⁸ R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).

⁹ J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, Science **264**, 553 (1994).

¹⁰ C.-J. Sheu and S.-L. Jang, Solid-State Electron. **44**, 1819 (2000).

¹¹ D. Vasileska and S. S. Ahmed, IEEE Trans. Electron Devices **52**, 227 (2005).

¹² S. Jin, M. V. Fischetti, and T. Tang, J. Appl. Phys. **102**, 083715 (2007).

¹³ E. B. Ramayya, D. Vasileska, S. M. Goodnick, and I. Knezevic, J. Appl. Phys. **104**, 063711 (2008).

¹⁴ R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. leBlanc, IEEE J. Solid-State Circuits **9**, 256 (1974).

- ¹⁵ J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze, *IEEE Electron Dev. Lett.* **1**, 2 (1980).
- ¹⁶ S. Takagi and A. Toriumi, *IEEE Trans. Electron Devices* **42**, 2125 (1995).
- ¹⁷ S. A. Hareland, S. Krishnamurthy, S. Jallepali, C.-F. Yeap, K. Hasnat, A. F. Tasch, Jr., and C. M. Maziar, *IEEE Trans. Electron Devices* **43**, 90 (1996).
- ¹⁸ D. Vasileska, D. K. Schroder, and D. K. Ferry, *IEEE Trans. Electron Devices* **44**, 584 (1997).
- ¹⁹ K. S. Krisch, J. D. Bude, and L. Manchanda, *IEEE Electron Dev. Lett.* **17**, 521 (1996).
- ²⁰ A. Buin, A. Verma, A. Svizhenko, and M. Anantram, *Nano Lett.* **8**, 760 (2008).
- ²¹ X. Lü, J. H. Chu, and W. Z. Shen, *J. Appl. Phys.* **93**, 1219 (2003).
- ²² X. Lü and J. Chu, *J. Appl. Phys.* **100**, 014305 (2006).
- ²³ Y. Chen, D. Li, J. R. Lukes, and A. Majumdar, *J. Heat Transfer* **127**, 1129 (2005).
- ²⁴ G. Bacarani and M. R. Worderman, in *Proceedings of the IEDM* p. 278 (1982).
- ²⁵ M.-S. Liang, J. Y. Choi, P.-K. Ko, and C. Hu, *IEEE Trans. Electron Devices* **33**, 409 (1986).
- ²⁶ A. Hartstein and N. F. Albert, *Phys. Rev. B* **38**, 1235 (1988).
- ²⁷ M. J. van Dort, P. H. Woerlee, A. J. Walker, C. A. H. Juffermans, and H. Lifka, *IEEE Trans. Electron Devices* **39**, 932 (1992).
- ²⁸ M. J. van Dort, P. H. Woerlee, and A. J. Walker, *Solid-State Electron.* **37**, 411 (1994).
- ²⁹ D. Vasileska and D. K. Ferry, in the *Technical Proceedings of the First International Conference on Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators* p. 408 (1998).
- ³⁰ <https://www.nanohub.org/tools/schred/>.
- ³¹ P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- ³² W. Kohn and L. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ³³ L. Hedin and S. Lindqvist, in *Solid State Physics*, edited by F. Seitz and D. T. and H. Ehrenreich (Academic Press, New York, 1969), vol. 23.
- ³⁴ M. Lundstrom, *Fundamentals of Carrier Transport* (Addison-Wesley Publishing Company Inc., New York, 1992).
- ³⁵ D. Vasileska, Ph.d. dissertation, Arizona State University (1995).
- ³⁶ P. J. Price, *Ann. Phys.* **133**, 217 (1981).
- ³⁷ B. K. Ridley, *Rep. Prog. Phys.* **54**, 169 (1991).
- ³⁸ D. Roychoudhury and P. K. Basu, *Rep. Prog. Phys.* **22**, 6325 (1980).
- ³⁹ C. Hao, J. Zimmermann, M. Charef, R. Fauquembergue, and E. Constant, *Solid-State Electron.* **28**, 733 (1985).
- ⁴⁰ S. M. Goodnick and P. Lugli, *Phys. Rev. B* **37**, 2578 (1988).
- ⁴¹ S. Imanaga and Y. Hayafuji, *J. Appl. Phys.* **70**, 1522 (1991).
- ⁴² W. Magnus, *Solid-State Electron.* **36**, 843 (1993).
- ⁴³ T. Yamada, J.-R. Zhou, H. Miyata, and D. K. Ferry, *IEEE Trans. Electron Devices* **41**, 1513 (1994).
- ⁴⁴ D. K. Ferry, *Phys. Rev. B* **14**, 5364 (1976).
- ⁴⁵ G. D. Mahan, *Many-Particle Physics* (Plenum, New York, 1981).
- ⁴⁶ J. M. Ziman, *Electrons and Phonons* (Oxford University Press, London, 1960).
- ⁴⁷ B. K. Ridley, *Quantum Processes in Semiconductors* (Oxford University Press Inc., New York, 1993).
- ⁴⁸ F. Seitz, *Phys. Rev.* **73**, 549 (1948).
- ⁴⁹ W. A. Harrison, *Phys. Rev.* **104**, 1281 (1956).
- ⁵⁰ C. Herring and E. Vogt, *Phys. Rev.* **101**, 944 (1956).
- ⁵¹ D. K. Ferry and C. Jacoboni, eds. (????).
- ⁵² H. J. G. Meyer, *Phys. Rev.* **112**, 298 (1958).
- ⁵³ H. A. G. de Cerdeira, *Solid State Comm.* **12**, 511 (1973).
- ⁵⁴ L. Pintschovius, J. A. Vergez, and M. Cardona, *Phys. Rev. B* **26**, 5658 (1982).
- ⁵⁵ D. K. Ferry, *Surf. Sci.* **57**, 218 (1976).
- ⁵⁶ D. K. Ferry, *Phys. Rev. B* **14**, 1605 (1976).
- ⁵⁷ S. Zollner, S. Gopalan, and M. Cardona, *J. Appl. Phys.* **68**, 1682 (1990).
- ⁵⁸ E. J. Moore, *Phys. Rev.* **160**, 607 (1967).
- ⁵⁹ W. Kohn and J. M. Luttinger, *Phys. Rev.* **108**, 590 (1957).
- ⁶⁰ J. M. Luttinger, in *Mathematical Methods in Solid State and Superfluid Theory* (Plenum Press, New York, 1967).
- ⁶¹ F. Stern and W. E. Howard, *Phys. Rev.* **163**, 816 (1967).
- ⁶² J. D. Jackson, *Classical Electrodynamics* (John Wiley and Sons, Inc., New York, 1975).
- ⁶³ J. J. Thomson, *Proc. Cambridge Philos. Soc.* **11**, 1120 (1901).
- ⁶⁴ K. Fuchs, *Proc. Cambridge Philos. Soc.* **34**, 100 (1938).
- ⁶⁵ E. H. Sondheimer, *Adv. Phys.* **1**, 1 (1952).
- ⁶⁶ E. Prange and T. Nee, *Phys. Rev.* **168**, 779 (1968).
- ⁶⁷ Z. Tesanovic, M. V. Jaric, and S. Maekawa, *Phys. Rev. Lett.* **57**, 2760 (1986).
- ⁶⁸ N. Trivedi and N. W. Ashcroft, *Phys. Rev. B* **38**, 12298 (1988).
- ⁶⁹ S. M. Goodnick, R. G. Gann, D. K. Ferry, and C. W. Wilmsen, *Surf. Sci.* **113**, 233 (1982).
- ⁷⁰ A. Gold, *Solid State Commun.* **60**, 531 (1986).
- ⁷¹ A. Gold, *Phys. Rev. B* **35**, 723 (1987).
- ⁷² G. Fishman and D. Calecki, *Phys. Rev. Lett.* **62**, 1302 (1989).
- ⁷³ H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka, and T. Matsusue, *Appl. Phys. Lett.* **51**, 1934 (1987).
- ⁷⁴ S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and L. Krivanek, *Phys. Rev. B* **32**, 8171 (1985).
- ⁷⁵ R. M. Feenstra, *Phys. Rev. Lett.* **72**, 2749 (1994).
- ⁷⁶ Y. Matsumoto and Y. Uemura, *Jpn. J. Appl. Phys. Suppl.* **2**, 367 (1974).
- ⁷⁷ T. Ando, *J. Phys. Soc. Japan* **43**, 1616 (1977).
- ⁷⁸ S. M. Goodnick and P. Lugli, *Phys. Rev. B* **52**, 5445 (1995).
- ⁷⁹ M. Mosko, A. Moskova, and V. Cambel, *Phys. Rev. B* **51**, 16860 (1995).
- ⁸⁰ D. Vasileska, P. Bordone, and D. K. Ferry, in *Proceedings of the Third International Workshop on Computational Electronics*, Portland, Oregon, May (1994).
- ⁸¹ N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (W. B. Saunders Co., New York, 1976).
- ⁸² E. D. Siggia and P. C. Kwok, *Phys. Rev. B* **2**, 1024 (1970).
- ⁸³ N. D. Mermin, *Phys. Rev. B* **1**, 2362 (1970).
- ⁸⁴ J. P. Walter and M. L. Cohen, *Phys. Rev. B* **5**, 3101 (1972).
- ⁸⁵ L. Wendler and R. Pechstedt, *Phys. Stat. Sol. B* **138**, 197 (1986).
- ⁸⁶ W.-Y. Chung and D. K. Ferry, *Solid-State Electron.* **31**, 1369 (1988).
- ⁸⁷ H. Haug and C. Ell, *Phys. Rev. B* **46**, 2126 (1992).
- ⁸⁸ K. S. Yi, A. M. Krizan, and D. K. Ferry, *Semicond. Sci. Technol.* **7**, B316 (1992).
- ⁸⁹ C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer-Verlag, Vi-

- enna, 1989).
- ⁹⁰ K. Hess, *Monte Carlo Device Simulation: Full Band and Beyond* (Kluwer Academic Publishing, Boston, 1991).
 - ⁹¹ M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986).
 - ⁹² S. Bosi and C. Jacoboni, J. Phys. C **9**, 315 (1976).
 - ⁹³ P. Lugli and D. K. Ferry, IEEE Trans. Electron Devices **32**, 2431 (1985).
 - ⁹⁴ H. Sakaki, Jpn. J. Appl. Phys. **19**, L735 (1980).
 - ⁹⁵ V. Arora, Phys. Stat. Sol. (b) **105**, 707 (1981).
 - ⁹⁶ J. Lee and H. Spector, J. Appl. Phys. **54**, 3921 (1983).
 - ⁹⁷ J. Lee and M. Vassell, J. Phys. C: Solid State Phys. **17**, 2525 (1984).
 - ⁹⁸ G. Sanders, C. Stanton, and Y. Yang, Phys. Rev. B **48**, 11067 (1993).
 - ⁹⁹ J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, IEEE Trans. Electron Devices **52**, 1589 (2005).
 - ¹⁰⁰ Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin, and G. Klimeck, IEEE Trans. Electron Devices **52**, 1097 (2005).
 - ¹⁰¹ C. Harris and E. O'Reilly, Physica E **32**, 341 (2006).
 - ¹⁰² M. Nolan, S.O'Callaghan, G. Fagas, and J. Greer, Nano Lett. **7**, 34 (2007).
 - ¹⁰³ K. Nehari, N. Cavassilas, and F. Michelini, Appl. Phys. Lett. **90**, 132112 (2007).
 - ¹⁰⁴ E. Gnani, S. Reggiani, A. Gnudi, P. Parruccini, R. Colle, M. Rudan, and G. Baccarani, IEEE Trans. Electron Devices **54**, 2243 (2007).
 - ¹⁰⁵ S.-M. Koo, A. Fujiwara, J.-P. Han, E. M. Vogel, C. A. Richter, and J. E. Bonevich, Nano Lett. **4**, 2197 (2004).
 - ¹⁰⁶ R. Kotlyar, B. Obradovic, P. Matagne, M. Stettler, and M. D. Giles, Appl. Phys. Lett. **84**, 5270 (2004).
 - ¹⁰⁷ J. Wang, E. Polizzi, A. Ghosh, S. Datta, and M. Lundstrom, Appl. Phys. Lett. **87**, 043101 (2005).
 - ¹⁰⁸ M. Lenzi, P. Palestri, E. Gnani, S. Reggiani, A. Gnudi, D. Esseni, L. Selmi, and G. Baccarani, IEEE Trans. Electron Devices **55**, 2086 (2008).
 - ¹⁰⁹ O. Gunawan, L. Sekaric, A. Majumdar, M. Rooks, J. Appenzeller, J. W. Sleight, S. Guha, and W. Haensch, Nano Lett. **8**, 1566 (2008).
 - ¹¹⁰ B. Auld, *Acoustic fields and waves in solids* (Wiley-Interscience, New York, 1973).
 - ¹¹¹ Y. S. Ju and K. E. Goodson, Appl. Phys. Lett. **74**, 3005 (1999).
 - ¹¹² E. P. Pokatilov, D. L. Nika, and A. A. Balandin, Phys. Rev. B **72**, 113311 (2005).
 - ¹¹³ E. P. Pokatilov, D. L. Nika, and A. A. Balandin, Superlattices and Microstructures **38**, 168 (2005).
 - ¹¹⁴ N. Nishiguchi, Y. Ando, and M. N. Wybourne, J. Phys.: Condens. Matter **9**, 5751 (1997).
 - ¹¹⁵ W. Weber, Phys. Rev. B **15**, 4789 (1977).
 - ¹¹⁶ S. P. Hepplestone and G. P. Srivastava, Phys. Status Solidi C **1**, 2617 (2004).
 - ¹¹⁷ R. W. Morse, J. Acoust. Soc. Am. **22**, 219 (1950).
 - ¹¹⁸ W. M. Visscher, A. Migliori, T. M. Bell, and R. A. Reinert, J. Acoust. Soc. Am. **90** (1991).
 - ¹¹⁹ L. Donetti, F. Gámiz, J. B. Roldán, and A. Godoy, J. Appl. Phys. **100**, 013701 (2006).
 - ¹²⁰ S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, Phys. Rev. B **32**, 8171 (1985).
 - ¹²¹ S. Takagi, J. Koga, and A. Toriumi, Jpn. J. Appl. Phys. **37**, 1289 (1998).
 - ¹²² K. Uchida, J. Koga, and S. Takagi, J. Appl. Phys. **102**, 074510 (2007).
 - ¹²³ E. B. Ramayya, D. Vasileska, S. M. Goodnick, and I. Knezevic, IEEE Trans. Nanotechnol. **6**, 113 (2007).
 - ¹²⁴ E. B. Ramayya, D. Vasileska, S. M. Goodnick, and I. Knezevic, J. Comput. Electron. **7**, 319 (2008).
 - ¹²⁵ <http://www.caam.rice.edu/software/ARPACK/>.
 - ¹²⁶ M. V. Fischetti and S. E. Laux, Phys. Rev. B **48**, 2244 (1993).
 - ¹²⁷ F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. E. Elewa, IEEE Electron Dev. Lett. **8**, 410 (1987).
 - ¹²⁸ F. Gamiz and M. V. Fischetti, J. Appl. Phys. **89**, 5478 (2001).
 - ¹²⁹ R. Shishir and D. K. Ferry, J. Comput. Electron. **7**, 305 (2008).
 - ¹³⁰ P. Lugli and D. Ferry, IEEE Trans. Electron Devices **ED-32**, 2431 (1985).